McGraw-Hill Encyclopedia of Science and Technology 11

McGraw-Hill Encyclopedia

McGRAW-HILL BOOK COMPANY INC.

NEW YORK CHICAGO SAN FRANCISCO DALLAS TORONTO LONDON

of Science and Technology

AN INTERNATIONAL REFERENCE WORK

IN FIFTEEN VOLUMES INCLUDING AN INDEX

VOLUME 11 PROJ RYE

(LEFT) The explosion of a dynamite cap at an exposure of 1/10 microsecond The explos on at the bottom of the picture represents the original position of the dynamite cap Particles of the metal case and particles from the end of the cap leave the V shaped shock waves revealed by the lighting *(photograph by H E Edgerton)* (RIGHT) Living eggs of *Ciona* in cleavage *(photograph by M V Edds Jr)*

McGRAW HILL ENCYCLOPEDIA OF SCIENCE AND TECHNOLOGY

Library of Congress Catalog Card Number 60 11000

Suggestions to the Readers

The basic plan of the Encyclopedia is explained here in order to facilitate its use.

The subject matter of the various disciplines or branches of science and technology is organized systematically: a general article provides a broad survey of the field, and a number of separate articles, alphabetically arranged, cover its main subdivisions and more specific aspects.

Cross references guide the reader from the general articles to the other articles into which the subject is subdivided, and from these to articles on more highly specialized phases of the subject. The cross references—there are about 50,000 of them—are printed in small capital letters so that they can be easily recognized. By means of the cross references a reader may find his way from ELECTRICAL ENGINEERING through ELECTRONICS and VACUUM TUBE to ELECTRON MOTION IN VACUUM or ELECTRON EMISSION. Or, following another line of cross references, the reader would be led to ELECTRIC POWER SYSTEMS, TRANSMISSION LINES, ELECTROMAGNETIC WAVE, and so on.

In general, each article begins with a definition of the title that states its scope and coverage. Usually only the scientific or technological sense is discussed. Most of the articles, after this statement, go on to increasingly complex and detailed considerations. A reader thus needs to proceed only as far as his inclinations and requirements dictate.

The Index, Volume 15, should be consulted to locate the discussion of topics covered in the Encyclopedia but not given in separate entries.

Every phylum, class, and order in the plant and animal kingdoms is allotted a separate article. Many of the more common families, genera, and species are covered either in one of the order articles or in a separate article under its own scientific or common name.

The adjectives electric and electrical are used in the following senses: Electric—containing, producing, arising from, actuated by, or carrying electricity, or capable of doing so, as, for instance, electric generator, electric motor, electric wiring. Electrical—related to, pertaining to, or associated with electricity, but not having its properties or characteristics, as, for example, electrical code, electrical engineering.

Words used as titles are, wherever possible, given in the singular to permit a consistent alphabetic arrangement. Titles are alphabetized by word and not by letter, for example:

Earth sciences
Earth tides
Earthmover
Earthquake

A word used as a noun precedes the same word used adjectivally, thus:

Mercury (element)
Mercury (planet)
Mercury battery

or

Circuit, electronic
Circuit breaker

Hyphened terms are alphabetized as single words, for example:

Animal virus
Animal-feed composition

Most of the longer articles contain bibliographies citing useful sources of further information. For additional bibliographical citations, the reader should refer to related articles (as indicated by the cross references in the article). Bibliographies are placed at the ends of articles or sometimes at the ends of major sections in long articles.

A list of initials and names of the contributors to the Encyclopedia is to be found in Volume 15. This list will permit quick identification of a contributor's initials after an article. Immediately following this list is a second list of encyclopedia contributors with their affiliations and the titles of articles each has written for the Encyclopedia.

McGraw-Hill Encyclopedia of Science and Technology

Projection systems, optical

Optical projection is the process whereby a real image of a suitably illuminated object is formed by an optical system in such a manner that it can be viewed photographed or otherwise observed Essential equipment in an optical projection system consists of a light source a condenser an object holder a projection lens and (usually) a screen on which the image is formed (Fig 1) For some important applications of optical projection see CINEMATOGRAPHY LANTERN SLIDES

The luminance of the image in the direction of observation will depend upon (1) the average luminance of the image of the light source as seen through the projection lens from the image point under consideration (2) the solid angle subtended by the exit pupil of the projection lens at this image point, and (3) the reflective or transmissive characteristics of the screen Usually it is desirable to have this luminance as high as possible Therefore with a given screen lens and projection d stan e h ha

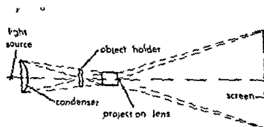


Fig 1 A simple optical projection system

The object is placed between the condenser and the projection lens If transparent it can be inserted directly in the light beam however it should be positioned and the optical system should be so designed that it does not vignette (cut off) any of the image of the light source in the projection lens
such
scope
tem
will
co
is ough a well corrected lens to a screen which is perpendicular to the optic axis of the len maximum illuminance of the image will occur on this axis and illuminance away from the axis will decrease in proportion to the fourth power of the cosine of the angle subtended with the axis at the projection lens In practice it is possible to design distortion into the condenser so that the illuminance is somewhat lower on the axis and considerably higher away from the axis than is given by this fourth power law Acceptable illumination for most visual purposes can allow a fall off from center to side in the image of as much as 50% particularly if the illuminance within a circle occupying one half of the image area does not drop below 80% of the maximum value

Light source Usually either an incandescent or an arc lamp is used as the light source To keep luminance high incandescent projection lamps operate at such high temperatures that their life is comparatively short Also they must be well cooled all except the smallest sizes require cooling fans for this purpose Filaments are finely coiled and accurately supported usually being carefully aligned in a prefocus base so that they will be precisely positioned in the optical system Spacing between coils is such that a small spherical mirror can be used back of the lamp to image the coils in the spaces between the coils thus increasing usable light output nearly twofold

When a highly uniform field is required a lamp consisting of a small disk of ceramic material heated to incandescence by radiofrequency induction is available With this it is possible to maintain illuminance of a projection field of several square inches with a variation of only 2 3% See INCANDESCENT LAMP

Arc lamps are used when incandescent lamps cannot provide sufficient flux to illuminate the screen area satisfactorily Carbon electrodes with

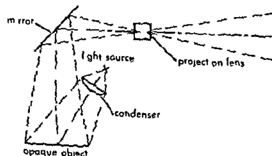


Fig 2 An epidiascope or system for projecting an image of an opaque object

special feed control magnetic arc stabilization and other devices are usually used to keep the arc as accurately positioned and as flicker free as possible A shield between the arc and the object called a douser is used to protect the object while the arc is ignited and to provide a quick shut off for the light beam Often water cells and other heat filtering devices are inserted in the beam to keep the heat on the object as low as possible *See* ARC LAMP

Condenser. The condenser system is used to gather as much of the light from the source as possible and to redirect it through the projection lens Both reflective and refractive systems are used Reflectors can be of aluminum although the better ones are of glass with aluminized coatings They are usually elliptical in shape with the light source at one focus and the image position in the projection lens at the other

Refractive systems may be of heat resistant glass or fused quartz. With arc lamps particularly the condenser lens is very close to the source in order to provide the high magnification required if the image is to fill the projection lens Usually therefore the condenser requires special cooling In larger projectors several elements are used to give the required magnification, these are often aspherical in shape to give the required light distribution

A well designed condenser can pick up light in a cone having a half angle in excess of 50° This means that with a well designed arc lamp or with an incandescent lamp using an auxiliary spherical mirror, more than one third of the total luminous flux radiated by the source can be directed through the projector

To obtain the high magnification required with arc sources and large aperture lenses a relay type of condenser illustrated in Fig 3 may be used This images the source beyond the first condenser system and then uses a second lens to relay this image to the projection lens This arrangement allows for better light distribution control in the screen image with less waste of light at the object position Also an Inconel "cat's eye" diaphragm at the first image point gives a convenient intensity control for the light beam For additional information on condensers *see* MICROSCOPE OPTICAL

Object holder The function of the object holder is to position the object exactly at the focal point of the projection lens Slight motion or vibration will reduce image sharpness Therefore proper mechanical design of this part of the system is most important When a succession of objects is to be projected the holder must be able to position these objects precisely and clamp them firmly in a minimum time Some designs have been suggested and a few tried which allow the object to move while the system projects a fixed image However, this requires a motion in some portion of the optical system which invariably reduces the quality of the projected image Because of this such systems have not found wide favor

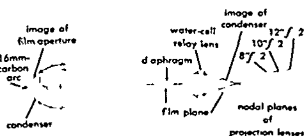


Fig 3 A relay condenser system having water cell incorporated in second stage

The object holder must also provide for protection of the object from unwanted portions of the beam for cooling the object where this may be required and for accurately adjusting its position laterally with respect to the beam In most systems focusing of the projection lens is provided by moving the lens itself rather than by disturbing the position of the object longitudinally in the system

Projection lens The function of the projection lens is to produce the image of the object on the screen Its design depends upon the use to be made of the projected image As examples a profile or contour projector requires a lens which is well corrected for the aberrations known as distortion and field curvature (*see* ABERRATION OPTICAL), an optical printer (or enlarger) requires a lens carefully designed to work at comparatively short conjugate focal distances and picture projectors must preserve the quality which has been captured by the camera lens

Since projection lenses usually transmit relatively intense light beams they must be designed to be heat resistant Their surfaces are usually not cemented Optical surfaces are coated to reduce reflection but these coatings must be able to withstand heat without deterioration Because camera lenses are not designed for this type of operation they should not be employed as projection lenses unless the projector is specifically designed to use them

Although an ideal position at which to control intensity is at the projector lens until 1955 it was not practical to incorporate a diaphragm in this lens In that year however large aperture lenses containing iris diaphragms were installed by Paramount Studios of Hollywood in their large process projectors and have been found to work very successfully (Fig 4)

Projection screen Usually a projection screen is used to redirect the light in the image for convenient observation Exceptions are systems which project the image directly into other optical systems for further processing or which form an aerial image which is to be viewed only from a localized position

Screens may be either reflective or translucent the latter type being used when the image is to be observed or photographed from the side away from the projector An example is the so-called self contained projector which has the screen on the housing holding the other elements Reflective screens may be matte having characteristics ap

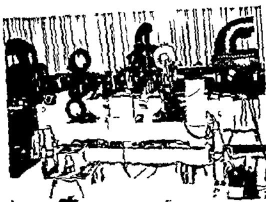


Fig 4 A triple-head projector using three 250-amp arc lamps and 12 in f/2 projection lenses. These lenses successfully incorporate iris diaphragms for intensity control (Paramount Pictures)

proaching Lambert reflection directional producing an image which will appear brighter in the direction of specular reflection or reflexive directing most of the light back toward the projector and giving an image of relatively high luminance with low projection intensity but with a very confined viewing angle [A J BU]

pr

E

form blackbody light source excited by radio frequency *J Soc Motion Picture Television Engrs.*, 64(12) 671–673 1955 M Reiss The cost law of illumination *J Opt Soc Am* 35(4) 283–288 1945

Projective geometry

A geometry that investigates those properties of figures that are unchanged (invariant) when the figures are projected from a point to a line or plane. Although isolated theorems appeared earlier the first systematic treatment of the subject was given by the French army officer J V Poncelet (1788–1867) who began his *Traité des propriétés projectives des figures* in 1812 while a prisoner of war in Russia. Two features of plane projective geometry are (1) introduction of an ideal line that each ordinary line g intersects (the intersection being common to all lines parallel to g) and (2) the principle of duality according to which any statement that is obtained from a valid one (theorem) by substituting for each concept involved its dual is also valid ("Line" and "point" are dual connecting two points by a line is dual to intersecting two lines and so on) The subject has been developed both synthetically (as a logical consequence of a set of postulates) and analytically (by the introduction of coordinates and the application of algebraic processes) The characteristic properties of geometrical projection are that it defines a one-to-one correspondence that preserves cross ratio between the two sets of entities involved (for example the points of two lines the lines of two pencils where a pencil of lines consists of all the lines on a point) Thus if P_i P'_i ($i = 1\ 2\ 3\ 4$) are corresponding points of two lines and x_i x'_i ($i = 1\ 2\ 3\ 4$) are their respective coordinates then

$$\frac{(x_3 - x_1)(x_4 - x_2)}{(x_3 - x_2)(x_4 - x_1)} = \frac{(x'_3 - x'_1)(x'_4 - x'_2)}{(x'_3 - x'_2)(x'_4 - x'_1)}$$

All such correspondences between one dimensional forms are given by the linear transformation $x' = (ax + b)/(cx + d)$ $ad - bc \neq 0$, between two-dimensional forms (for example two planes) by

$$x' = (a_{11}x + a_{12}y + a_{13})/(a_{31}x + a_{32}y + a_{33})$$
$$y' = (a_{21}x + a_{22}y + a_{23})/(a_{31}x + a_{32}y + a_{33})$$

with nonvanishing determinant $|a_{ij}|$ ($i\,j = 1\ 2\ 3$) The latter transformations form an eight parameter group the projective group of the plane It contains as a proper subgroup the three parameter group (rotations and translations) that defines euclidean geometry of the plane and so euclidean geometry is a subgeometry of projective geometry. *Conics are defined as the class of points of intersection* of corresponding lines of two projective pencils of lines (on distinct points) If these intersections are collinear the *pencils are perspective* and the conic is degenerate Otherwise the conic is nondegenerate Any two *nondegenerate conics* are equivalent that is a projective transformation exists that carries one into the other One of the well known theorems of projective geometry was proved by the French architect G Desargues (1593 1662) *If the lines joining corresponding vertices of two triangles are concurrent the corresponding sides intersect in collinear points* The converse of the theorem is its dual and hence is also valid *See* CONFORMAL MAPPING GEOMETRY EUCLIDEAN [L M BL]

Projector (light)

A device designed to produce controlled beams of light that can be projected over considerable distances The function of a light projector may *be to light a limited area* such as the face of a building or an actor on a stage or to produce apparent brightnesses at the light source that make the source itself visible from considerable distances

Equipment for the projection of light beams usually makes use of one or more of three *principles* of light control

1 *A specular* paraboloidal reflector (Fig 1) produces parallel rays of light from a source of brightness located at the focal point of the parabola Since practical light sources have finite dimensions some divergence from a parallel beam is usually produced Also if the light source is moved away from the focal point, a spreading of the beam is produced

2 *A* specular ellipsoidal reflector (Fig 2) with a light source at the near focal point focuses rays at the opposite focal point of the ellipse Again with sources of finite dimensions true point focusing is not accomplished but the rays are concen

trated in a small area in the region of the second focal point

3 A lens refracts the rays from a light source located at its focal point into parallel rays (see Fig 3)

Searchlights employ the first of these principles using carefully formed paraboloidal reflectors The light source, either a carbon arc or a compact incandescent filament is for practical purposes at the focal point Light from the filament or arc that would normally be scattered widely through the front of the searchlight is redirected through the focus by a spherical reflector The result is to produce a beam of maximum practical intensity and minimum divergence, capable of producing sufficient illumination on an object thousands of feet away to render it visible at night

Spotlights used in theatrical productions and in some forms of display lighting usually use filament lamps and lenses or ellipsoidal reflectors and lenses Ellipsoidal reflector spotlights have a lens focused at or near the second focal point of the ellipse to produce a nearly parallel beam from the light concentrated there Supplementary apertures may be introduced at the second focus to shape the beam into a desired shape or to reduce the divergence of the beam

Floodlights, used for outdoor lighting of buildings, parking lots sports fields and the like, usually use filament or mercury vapor lamps in conjunction with parabolic reflectors Where nonparallel beams are desired the reflector may be etched to scatter the light slightly or prismatic cover glasses may be employed to spread the beam by refracting some of the rays away from the beam axis Floodlights using fluorescent lamps are also used commercially but their beams are not usually precisely limited Economical reflector sizes are quite small with respect to the size of the fluorescent tube which does not therefore, produce the effects of a point source In applications where closely con

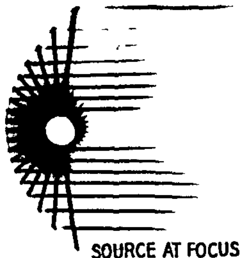


Fig 1 Paraboloidal reflector

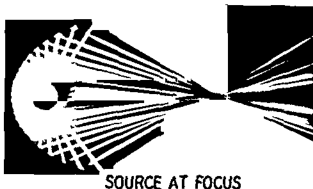


Fig 2 Ellipsoidal reflector

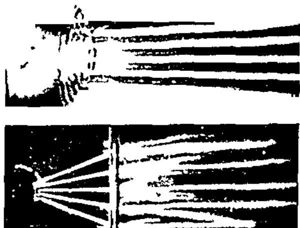


Fig 3 Lenses

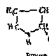
fined beams are not essential however the high efficiency of fluorescent lamps makes fluorescent floodlighting quite practical

Light beacons make use of the visibility of high intensity light sources at great distances Beacons were first used in lighthouses which identified landmarks for ships approaching shore Beacons have found even greater application in locating airports for aircraft at night Airway beacons usually employ filament lamps and lenses to provide a high intensity beam of narrow horizontal divergence The beam has a slight vertical divergence so that it covers an altitude of several thousand feet at a range of several miles As the beam rotates the narrow horizontal spread produces a flashing effect from the pilot's viewpoint This flashing effect tends to increase the beacon's visibility helping him to prepare for his approach as early as possible *See* ILLUMINATION [A MA]

Prolamine

A general name for a group of proteins soluble in 70 80% ethanol but insoluble in absolute ethanol water and other neutral solvents Their unusual solubility properties are attributable to their low content of polar side chains They are found primarily in plant products for example zein from corn gliaden from wheat hordein from barley *See* BARLEY, CORN, PROTEIN, WHEAT [D M T]

Proline



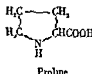
Physical constants of the L-isomer at 25 C
pK₁ COOH 1 99 pK₂ NH 10 96
Isoelectric point 6 30
Optical rotation [α] H₂O −86.2 [α]ᴅ 5 N HCl −60 4
Solubility (g/100 ml H₂O) very soluble

An amino acid The amino acids are characterized physically by the following (1) the pK_a, or the dissociation constant of the various titratable groups (2) the isoelectric point or pH at which a dipolar ion does not migrate in an electric field, (3) the optical rotation or the rotation imparted to a beam of plane polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml, (4) *solubility* See EQUILIBRIUM IONIC, ISOELECTRIC POINT OPTICAL ACTIVITY, SPECTROPHOTOMETRIC ANALYSIS

Proline forms a yellow color with ninhydrin reagent on paper chromatograms and is the precursor of hydroxyproline a major constituent of collagen (see CHROMATOGRAPHY) Proline is formed, biosynthetically, from glutamic acid (*see* AMINO ACIDS)

During metabolic degradation, oxidation of proline can take place in two ways The amino acid oxidases produce Δ' pyrroline 2 carboxylic acid

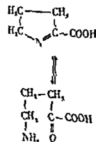
librium with glutamic acid ... PHOSPHOPYRIDINE NUCLEOTIDE (DPN)

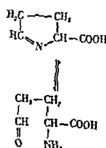


Proline

Oxidase

Dehydrogenase





α Keto-δ-aminovaleric acid

Glutamic acid semialdehyde

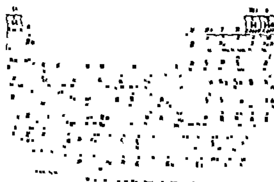
Glutamic acid semialdehyde is oxidized to glutamic acid as illustrated

[E A AD]

Promethium

Element number 61, promethium Pm is the "missing" *element* of the rare-earth series The most abundant separated isotope has the atomic weight 147

Although a number of scientists have claimed to have discovered this element in nature as a result of observing certain spectral lines no one has succeeded in isolating element 61 from naturally occurring materials It has been produced artificially in nuclear reactors since it is one of the products that result from the fission of uranium, thorium and plutonium In 1945 J A Marinsky, L. E. Glendenin and C Coryell isolated element 61 from fission product residues The chemical and *metallurgical* properties of promethium are very similar to those of neodymium and samarium Facilities have been set up to separate Pm^{147} at the



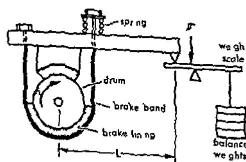
Oak Ridge National Laboratory and by April 1958 2 grams had been recovered All the known isotopes are radioactive and Pm^{147}, with a half life of 2 66 years is the isotope which is most commonly isolated Its principal uses are for research involving tracers ... industry
$Pm(NO_3$ ...

[F H SP]

Prony brake

A ...

... surface The prony



Prony brake (*From T Baumeister ed*, *Mechanical Engineers Handbook 6th ed McGraw Hill 1958*)

trated in a small area in the region of the second focal point.

3. A lens refracts the rays from a light source located at its focal point into parallel rays (see Fig. 3).

Searchlights employ the first of these principles, using carefully formed paraboloidal reflectors. The light source, either a carbon arc or a compact incandescent filament, is for practical purposes at the focal point. Light from the filament or arc that would normally be scattered widely through the front of the searchlight is redirected through the focus by a spherical reflector. The result is to produce a beam of maximum practical intensity and minimum divergence, capable of producing sufficient illumination on an object thousands of feet away to render it visible at night.

Spotlights used in theatrical productions and in some forms of display lighting usually use filament lamps and lenses or ellipsoidal reflectors and lenses. Ellipsoidal reflector spotlights have a lens focused at or near the second focal point of the ellipse to produce a nearly parallel beam from the light concentrated there. Supplementary apertures may be introduced at the second focus to shape the beam into a desired shape or to reduce the divergence of the beam.

Floodlights, used for outdoor lighting of buildings, parking lots, sports fields and the like, usually use filament or mercury vapor lamps in conjunction with parabolic reflectors. Where nonparallel beams are desired, the reflector may be etched to scatter the light slightly or prismatic cover glasses may be employed to spread the beam by refracting some of the rays away from the beam axis. Floodlights using fluorescent lamps are also used commercially but their beams are not usually precisely limited. Economical reflector sizes are quite small with respect to the size of the fluorescent tube, which does not, therefore, produce the effects of a point source. In applications where closely con-

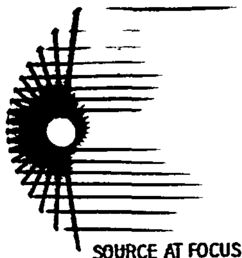


Fig. 1 Paraboloidal reflector

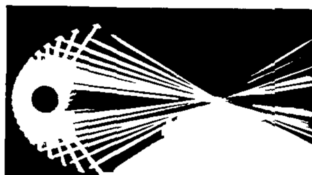


Fig. 2 Ellipsoidal reflector

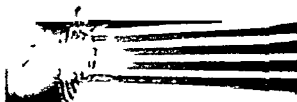




Fig. 3 Lenses

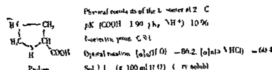
fined beams are not essential, however, the high efficiency of fluorescent lamps makes fluorescent floodlighting quite practical.

Light beacons make use of the visibility of high intensity light sources at great distances. Beacons were first used in lighthouses, which identified landmarks for ships approaching shore. Beacons have found even greater application in locating airports for aircraft at night. Airway beacons usually employ filament lamps and lenses to provide a high intensity beam of narrow horizontal divergence. The beam has a slight vertical divergence, so that it covers an altitude of several thousand feet at a range of several miles. As the beam rotates, the narrow horizontal spread produces a flashing effect from the pilot's viewpoint. This flashing effect tends to increase the beacon's visibility, helping him to prepare for his approach as early as possible. See ILLUMINATION. [A.M.A.]

Prolamine

A general name for a group of proteins soluble in 70–80% ethanol but insoluble in absolute ethanol, water, and other neutral solvents. Their unusual solubility properties are attributable to their low content of polar side chains. They are found primarily in plant products, for example, zein from corn, gliadin from wheat, hordein from barley. See BARLEY; CORN; PROTEIN; WHEAT. [D.S.T.]

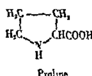
Proline



An amino acid The amino acids are characterized physically by the following (1) the pK_1 or the dissociation constant of the various titratable groups, (2) the isoelectric point or pH at which a dipolar ion does not migrate in an electric field, (3) the optical rotation or the rotation imparted to a beam of plane polarized light (frequently the D line of the sodium spectrum) passing through 1 decimeter of a solution of 100 grams in 100 ml, (4) solubility See EQUILIBRIUM, IONIC, ISOELECTRIC POINT OPTICAL ACTIVITY, SPECTROPHOTOMETRIC ANALYSIS

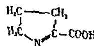
Proline forms a yellow color with ninhydrin reagent on paper chromatograms and is the precursor of hydroxyproline a major constituent of collagen (see CHROMATOGRAPHY) Proline is formed biosynthetically from glutamic acid (see AMINO ACIDS)

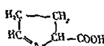
During metabolic degradation oxidation of proline can take place in two ways The amino acid oxidases produce Δ' pyrroline 2 carboxylic acid which is in equilibrium with α keto δ aminovaleric acid A diphosphopyridine nucleotide (DPN) linked dehydrogenase oxidizes proline to Δ' pyrroline 5 ... qui ... Di

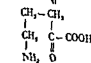


Proline

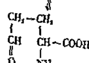








α keto-δ-aminovaleric acid



Glutamic acid semialdehyde

Glutamic acid semialdehyde is oxidized to glutamic acid as illustrated

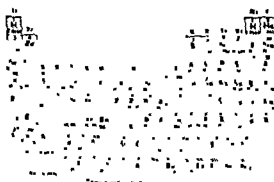
[E A AD]

Promethium

Element number 61, promethium Pm is the "missing" element of the rare-earth series The most abundant separated isotope has the atomic weight 147

Although a number of scientists have claimed to have discovered this element in nature as a result of observing certain spectral lines no one has succeeded in isolating element 61 from naturally occurring materials It has been produced artificially in nuclear reactors, since it is one of the products that result from the fission of uranium thorium and plutonium In 1945 J A Marinsky, L E Glendenin and C Coryell isolated element ...

cilities have been set up to separate Pm^{147} at the

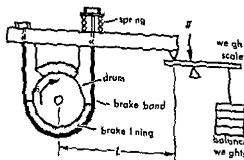


Oak Ridge National Laboratory and by April 1958 2 grams had been recovered All the known isotopes are radioactive, and Pm^{147}, with a half life of 2 66 years is the isotope which is most commonly isolated Its principal uses are for research involving tracers Its main application is in the phosphor industry The oxide Pm_2O_3 is purple, the nitrate $Pm(NO_3)_3$ is pink See RARE EARTH ELEMENTS

[E H SP]

Prony brake

... surface The prony



Prony brake (From T Baumeister, ed , Mechanical Engineers' Handbook 6th ed , McGraw Hill, 1958)

brake provides means for measuring the torque T developed as illustrated

$$T = L(W - W_0)$$

where L is the distance shown on the drawing W is the scale weight with the brake operating and W_0 is the scale weight with the brake free When the shaft speed as measured by a tachometer is n rpm brake horsepower is $2\pi nL(W - W_0)/33\,000$ Small prony brakes are air cooled the drum of larger ones may be filled with water to absorb the heat during operation [F H R]

Propane

A member of the alkane or paraffin series of hydrocarbons formula $CH_3CH_2CH_3$ It makes up 3–18% of natural gas It is readily liquefied (melting point −187 7°C boiling point −42 1°C) and mixtures with liquefied butane are sold as liquefied petroleum gas (LPG) in cylinders under moderate pressure for domestic fuel

At temperature above about 650°C propane undergoes cracking to ethylene and methane This reaction is the basis of an important commercial source of ethylene The reaction is accompanied by some dehydrogenation to propylene The yield of propylene is increased in the presence of catalyst

In the petroleum industry propane is used as a combined solvent and refrigerant for the refining of lubricants and other products See ALKANE CRACKING PETROLEUM PROCESSING [L S]

Propanol

One of the three carbon saturated aliphatic alcohols Normal propyl alcohol $CH_3CH_2CH_2OH$ also known as 1 propanol or ethyl carbinol is very similar to ethanol in its chemistry and properties It is a colorless mobile and toxic liquid (molecular weight 60 09 melting point −127°C boiling point 97 15°C specific gravity 0 8036) with a pungent odor It is soluble in water and most organic liquids

n Propanol is produced in small amounts as a by product in the synthesis of methanol from carbon monoxide and hydrogen in the oxidation of propane and butane and in the Fischer Tropsch reaction The oxo synthesis from ethylene carbon monoxide and hydrogen is an attractive route for large-scale manufacture of the compound Its major uses are as a solvent and chemical intermediate

Isopropyl alcohol $CH_3CHOHCH_3$ also known as 2 propanol and dimethyl carbinol is the simplest of the secondary alcohols and is a major industrial organic chemical over 1 000 000 000 lb were produced in 1956 in the United States alone It is a colorless mobile and toxic liquid (molecular weight 60 09 melting point −89 5°C boiling point 82 4°C specific gravity 0 78(3) which has a pungent odor and taste and is soluble in water and most organic liquid

It is produced by the hydration of propylene with sulfuric acid and water and is used mainly for the manufacture of acetone by catalytic dehydrogenation Other uses are as a solvent extractant antifreeze and rubbing alcohol See ALCOHOL [J W L]

Propellant

Usually a combustible substance that produces heat and supplies ejection particles as in a rocket engine A propellant is both a source of energy and a working substance a fuel is chiefly a source of energy and a working substance is chiefly a means for expending energy (see AIRCRAFT FUEL FUEL THERMODYNAMIC CYCLE) Because the distinction is more decisive in rocket engines the term propellant is used primarily to describe chemicals carried by rockets for propulsive purposes

Propellants are classified as liquid or as solid Even if a propellant is burned as a gas it may be carried under pressure as a cryogenic liquid to save space For example liquid oxygen and liquid hydrogen are important high energy liquid bipropellants Liquid propellants are carried in compartments separate from the combustion chamber solid propellants are carried in the combustion chamber The two types of propellant lead to significant differences in engine structure and thrust control (see ROCKET ENGINE) For comparison the effectiveness of either type of propellant is stated in terms of specific impulse See SPECIFIC IMPULSE THRUST

LIQUID PROPELLANT

A liquid propellant releases energy by chemical action to supply motive power for jet propulsion The three principal types of propellants are monopropellant bipropellant and hybrid propellant Monopropellants are single liquids either compounds or solutions Bipropellants consist of fuel and oxidizer carried separately in the vehicle and brought together in the engine Air breathing engines carry only fuel and use atmospheric oxygen for combustion Hybrid propellants use a combina

Physical properties of liquid propellants

Propellant	Boiling point °F	Freezing point °F	Density g ml	Specific impulse* sec
Monopropellants				
Acetylene	119	115	0 62	265
Hydrazine	236	35	1 01	194
Ethylene oxide	52	−168	0 88	192
Hydrogen peroxide	288	13	1 39	170
B propellants				
Hydrogen	423	436	0 07	
Hydrogen fluorine	306	360	1 54	410
Hydrogen-oxygen	−297	362	1 14	390
Nitrogen tetroxide	70	12	1 49	
Nitrogen tetroxide hydrazine	236	3	1 01	290
Red nitric acid	104	80	1 58	
Red fuming nitric acid unsym methyl hydrazine	146	−71	0 78	2[illegible]

* Maximum theoretical specific impulse at 1000 psi chamber pressure expanded to atmospheric pressure

tion of liquid and solid materials to provide propulsion energy and working substance (*see* METAL BASE FUEL) Typical liquid propellants are listed in the table Physical properties at temperatures from storage to combustion are important These properties include melting point boiling point density and viscosity

Performance Jet propulsion by a reaction engine using the momentum of the propellant combustion products ejected from the engine is not limited to atmospheric operation if the fuel reacts with an oxidizer carried with the engine Performance of the propellant in such an engine depends upon both the heat liberated and the propellant reaction products Combustion with air of effective fuels for air breathing engines gives approximately 18 000 Btu/lb whereas fuels which are more effective in rocket engines may give only 15 000 Btu/lb A high heat of reaction is most effective with gaseous products of low molecular weight

Performance is rated in terms of specific impulse (occasionally specific thrust) the thrust obtained per pound of propellant used in 1 sec An alternate measure of performance is the characteristic exhaust velocity The theoretical characteristic exhaust velocity is determined by the thermodynamic properties of the propellant reaction and its products Unlike the specific impulse the characteristic exhaust velocity is independent of pressure except for second order effects such as reactions modifying the heat capacity ratio of the combustion gases

These parameters are related as shown in the equation

$$I_s = F/w = c^* C_F / g$$

where I_s is specific impulse in seconds F is thrust and w is flow rate of propellant The characteristic exhaust velocity c^* is given in feet per second C_F is the thrust coefficient and g is the gravitational constant

The actual exhaust velocity of the combustion gases is given by the product of the characteristic exhaust velocity and the thrust coefficient $c^* C_F$ The thrust coefficient is a function of the heat capacity ratio of the combustion gases (the ratio of the heat capacity at constant pressure to the heat capacity at constant volume) and of the ratio of the chamber pressure to the exhaust pressure

The permit heat capacity of common propulsion gases varies from 1 1 to 1 4 Increase in combustion chamber pressure increases specific impulse (Fig 1) Variation in the stoichiometry of the propellant reaction (the oxidizer fuel ratio) also affects performance A slightly fuel rich reaction gives higher performance with common liquid propellants despite the lower heat of reaction because of more favorable working gas composition Increase in chamber pressure usually moves the optimum performance point toward the stoichiometric reaction ratio A nonstoichiometric ratio may be used to give low combustion temperatures if required by structural materials

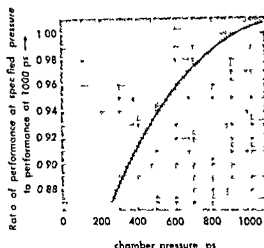


Fig 1 Approximate effect of chamber pressure on specific impulse

A properly designed engine can give 95–100% of the theoretical performance shown in the table

Combustion The energy of liquid propellants is released in combustion reactions which also produce the working fluid for reaction propulsion The liquids in a bipropellant system may ignite spontaneously on contact (hypergolic liquids) or they may require an ignition device to raise them to the flash or ignition temperature (anergolic liquids) Combustion can be initiated with a spark a hot wire or an auxiliary hypergolic liquid Monopropellant combustion or more properly decomposition can also be ignited by catalysis with an active surface or by a chemical compound in solution Ignition of common hypergolic bipropellants occurs in a period of 1–100 milliseconds following initial contact of the liquids Catalytic quantities of detergents or of certain compounds of metals with several oxidation states such as vanadium pentoxide decrease the ignition delay period of specific bipropellants

The combustion chamber in operation contains a turbulent heterogeneous high temperature reaction mixture The liquids burn with droplets of various sizes in close proximity and traveling at high velocity Larger masses of liquid may be present particularly at the chamber walls Very high rates of heat release of the order of 10^5–10^6 Btu/(min)(ft³)(atm) are encountered

Oscillations with frequencies of 25–10 000 cps or more may accompany combustion of liquids in jet propulsion engines Low frequency instability (chugging) can result from oscillations coupling the liquid flowing into the combustion chamber with pressure pulses in the chamber Higher frequencies (screaming) can result from gas oscillations of the acoustic type in the chamber itself

Engine performance in contrast to theoretical propellant performance depends upon effective combustor design Mixing and atomization are essential factors in injection of the propellants into the combustion chamber Injector and chamber

brake provides means for measuring the torque T developed as illustrated

$$T = L(W - W_0)$$

where L is the distance shown on the drawing W is the scale weight with the brake operating and W_0 is the scale weight with the brake free When the shaft speed as measured by a tachometer is n rpm brake horsepower is $2\pi nL(W - W_0)/33\,000$ Small prony brakes are air cooled the drum of larger ones may be filled with water to absorb the heat during operation [F H R]

Propane

A member of the alkane or paraffin series of hydro carbons formula $CH_3CH_2CH_3$ It makes up 3–18% of natural gas It is readily liquefied (melting point −187 7°C boiling point −42 1°C) and mixtures with liquefied butane are sold as liquefied petroleum gas (LPG) in cylinders under moderate pressure for domestic fuel

At temperature above about 650°C propane un dergoes cracking to ethylene and methane This reaction is the basis of an important commercial source of ethylene The reaction is accompanied by some dehydrogenation to propylene The yield of propylene is increased in the presence of catalyst

In the petroleum industry propane is used as a combined solvent and refrigerant for the refining of lubricants and other products *See* ALKANE CRACKING PETROLEUM PROCESSING [L S]

Propanol

One of the three carbon saturated aliphatic alcohols Normal propyl alcohol CH CH CH_2OH also known as 1 propanol or ethyl carbinol is very similar to ethanol in its chemistry and properties It is a colorless mobile and toxic liquid (molecu lar weight 60 09 melting point −127°C boiling point 97 15°C specific gravity 0 8036) with a pungent odor It is soluble in water and most or ganic liquids

n Propanol is produced in small amounts as a by product in the synthesis of methanol from car bon monoxide and hydrogen in the oxidation of propane and butane and in the Fischer Tropsch reaction The oxo synthesis from ethylene carbon monoxide and hydrogen is an attractive route for large scale manufacture of the compound Its ma jor uses are as a solvent and chemical intermedi ate

Isopropyl alcohol $CH_3CHOHCH_3$ also known as 2 propanol and dimethyl carbinol is the simplest of the secondary alcohols and is a major industrial organic chemical over 1 000 000 000 lb were pro duced in 1956 in the United States alone It is a colorless mobile and toxic liquid (molecular weight 60 09 melting point −89 5°C boiling point 82 4°C specific gravity 0 7863) which has a pungent odor and taste and is soluble in water and most organic liquids

It is produced by the hydration of propylene with sulfuric acid and water and is used mainly for the manufacture of acetone by catalytic dehydro genation Other uses are as a solvent extractant antifreeze and rubbing alcohol *See* ALCOHOL [J W L]

Propellant

Usually a combustible substance that produces heat and supplies ejection particles as in a rocket engine A propellant is both a source of energy and a working substance a fuel is chiefly a source of energy and a working substance is chiefly a means for expending energy (*see* AIRCRAFT FUEL FUEL THERMODYNAMIC CYCLE) Because the distinction is more decisive in rocket engines the term propel lant is used primarily to describe chemicals carried by rockets for propulsive purposes

Propellants are classified as liquid or as solid Even if a propellant is burned as a gas it may be carried under pressure as a cryogenic liquid to save space For example liquid oxygen and liquid hydrogen are important high energy liquid bipro pellants Liquid propellants are carried in compart ments separate from the combustion chamber solid propellants are carried in the combustion chamber The two types of propellant lead to signi ficant differences in engine structure and thrust control (*see* ROCKET ENGINE) For comparison the effectiveness of either type of propellant is stated in terms of specific impulse *See* SPECIFIC IMPULSE THRUST

LIQUID PROPELLANT

A liquid propellant releases energy by chemical *action to supply motive power for jet propulsion* The three principal types of propellants are mono propellant bipropellant and hybrid propellant Monopropellants are single liquids either com pounds or solutions Bipropellants consist of fuel and oxidizer carried separately in the vehicle and brought together in the engine Air breathing en gines carry only fuel and use atmospheric oxygen for combustion Hybrid propellants use a combina

Physical properties of liquid propellants

Propellant	Boiling point °F	Freezing point °F	Density g ml	Specific impulse * sec
Monopropellants				
Acetylene	−119	115	0 62	265
Hydrazine	236	35	1 01	194
Ethylene oxide	52	168	0 88	197
Hydrogen peroxide	288	13	1 39	170
B propellants				
Hydrogen	423	436	0 07	
Hydrogen fluorine	306	360	1 74	410
Hydrogen-oxygen	297	362	1 14	390
Nitrogen tetroxide	70	12	1 49	
Nitrogen tetroxide-hydrazine	236	3	1 01	290
Red nitric acid	104	80	1 58	
Red fuming nitric acid uns-dimethyl hydrazine	146	−71	0 78	27

* Maximum theoretical specific impulse at 1000 psi chamber pressure expanded to atmospheric pressure

Fig 3 Typ cal thrust t me character st cs for propellant w th a cyl ndrical cav ty

most any need (Fig 3) Another important neutral grain design the uninhibited internal external burning cylinder is used widely in short-duration appl cations such as the bazooka rocket weapon which contains many such grains Obturation is not necessary because the propellant is burned so rap dly that the walls of the chamber do not rise in temperature sufficiently to cause loss of strength

Propellant burns at a rate r proportional to a fractional power n of the pressure P as expressed by the equation

$$r = K_1 P^n$$

where K_1 is the coefficient of proportionality This rate may be determined at various pressures by measurements of the burning rate of propellant strands in a strand burner (Fig 4) If the propel lant is to operate properly the exponent n in the burning rate equation must be less than 1 As illustrated by Fig 5a if $n < 1$ there is a stable operating pressure at which the lines of fluid gen eration and fluid discharge intersect Pressure can not rise above this value because gas would then be discharged at a faster rate than it is generated by burning of propellant When the propellant meets the requirement $n < 1$ a relationship exists be tween operating chamber pressure and a design parameter known as K_n which is the ratio of pro pellant burning area to nozzle throat area as shown in Fig 5b

Specific impulse of solid propellants is normally rated with the rocket operating at chamber pres sure of 1000 lb and exhausting through an optimum nozzle into sea level atmosphere Under these con d t ons sol d propellants in use today can give an impulse of about 250 which is near the ceiling imposed by compositions based on ammonium per chlorate and hydrocarbons and is 5–10% lower than impulses obtainable from liquid oxygen and gasoline

f 1 d n opellants is

solid propellant rockets are easy to launch are instantly ready and have demonstrated a high degree of reliability Because they can be produced by a process much like the casting of concrete

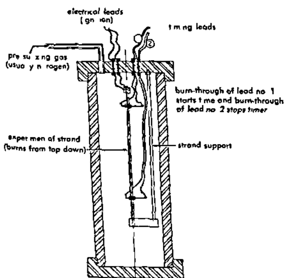


Fig 4 Strand burner apparatus

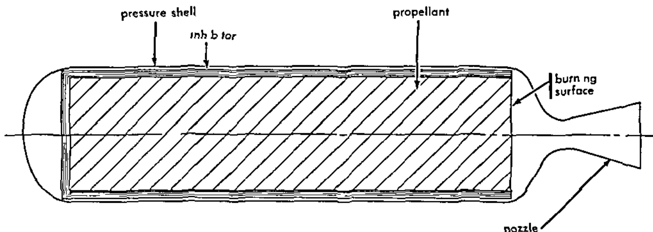


Fig 2 End burn ng grain loaded in rocket

design influence the flow pattern of both liquid and gases in the chamber The characteristic chamber length is given by

$$L^* = V_c / A_T$$

where V_c is the chamber volume and A_T is the area of the nozzle throat In general monopropellants require larger L^* than bipropellants to provide an equal fraction of theoretical performance in a rocket engine as expected from the slower combus tion exhibited by monopropellants [s si]

SOLID PROPELLANT

A solid propellant is a mixture of oxidizing and reducing materials that can coexist in the solid state at ordinary temperatures When ignited a propellant burns and generates hot gas Although gun powders are sometimes called propellants the term solid propellant ordinarily refers to materials used to furnish energy for rocket propulsion

Composition *A solid propellant normally con* tains three essential components oxidizer fuel and additives Oxidizers commonly used in solid propellants are ammonium and potassium perchlo rates ammonium and potassium nitrates and vari ous organic nitrates such as glyceryl trinitrate (nitroglycerin) Common fuels are hydrocarbons or hydrocarbon derivatives such as synthetic rub bers synthetic resins and cellulose or cellulose *derivatives The additives usually present in* small amounts are chosen from a wide variety of materi als and serve a variety of purposes Catalysts or suppressors are used to increase or decrease the rate of burning ballistic modifiers may be used for a variety of reasons such as providing less change in burning rate with pressure (platinizing agent) stabilizers may be used to slow down un desirable changes that may occur in long term stor age

Solid propellants are classified as composite or double base The composite types consist of an oxidizer of inorganic salt in a matrix of organic fuels such as ammonium perchlorate suspended in a synthetic rubber The double-base types are usually high strength high modulus gels of cellu lose nitrate (guncotton) in glyceryl trinitrate or a similar solvent

Propellants are processed by extrusion or casting techniques into what are often intricate shapes that are commonly called grains even though they may weigh many tons The double base types and certain high modulus composites are processed into grains by casting or extrusion and are then loaded by insertion of the cartridgelike grain or grains into the combustion chamber of the rocket This technique requires some type of mechanical sup port to hold the propellant in place in the chamber Certain types of composite propellants bonded by elastomeric fuels can be cast directly into the chamber where the binder cures to a rubber and the grain is then supported by adhesion to the chamber walls Most high performance rockets are made by this case bonding technique because more efficient use can be made of weight and combustion chamber volume

Burning rate The thrust time characteristic of a solid propellant rocket is controlled by geometric *shape of the grain Often it is desired that burning* not take place on certain portions of the grain Such surfaces are then covered with an inert ma terial called an inhibitor or restrictor Neutral burning grains maintain a constant surface during burning and produce a constant thrust Progressive-burning grains increase in surface and give an increasing thrust with time Degressive or regres sive grains burn with decreasing surface and give a decreasing thrust

An end burning grain is shown in Fig 2 This type of configuration is neutral because the surface stays constant while the grain burns forward For most applications radial burning charges which burn outward from the inside perforation are superior because most of the wall area of the cham ber can be protected from hot gas generated by combustion Such protection is a built in feature of the case bonded grain with the cartridge loaded inhibited charge protection is provided by the addi tion of obturators to prevent gas flow around the outside of the grain

Figure 2 shows a progressive design called an internal burning cylinder Various star shaped per forations can be used to give neutral or degressive characteristics By ingenious use of geometry the thrust time characteristic can be designed to meet

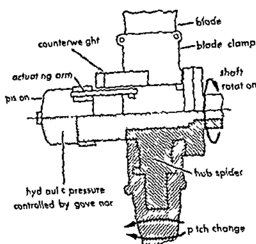


Fig 2 Actuating mechanism for changing propeller pitch

tating the blades in the hubs are ordinarily classified as hydraulic electric mechanical and automat

linked to the root of the blades for rotational movement (Fig 2) Oil may be regulated either manually as in the case of the two-position or variable pitch propellers or by means of a governor in the case of the constant speed propeller In the electrically actuated mechanism an electric motor is geared to the blade roots for slow rotational movement This motor may be regulated manually or by a governor Mechanically actuated pitch mechanisms are generally confined to small low powered aircraft in which the pitch change forces are small enough to be handled by manual means Automatic pitch *propellers have been* used to a lim ed

... to decrease pitch and centrifugal force to increase pitch torque to decrease pitch and centrifugal force to increase pitch and thrust to decrease pitch and a spring to increase pitch

Propellers designed for gas turbines require special features to control pitch *because the turbine* is operated at high rotational speed at all times even during taxiing on the ground In contrast the piston engine speed is reduced to an idle during landing and taxiing The problem of propeller structure is simplified for the gas turbine application however because of the absence of high vibrational exciting forces caused *by the engine*

Construction The material used in the blades of early or ...

... for many ... for propeller diameters of over 13 ft hollow blade construction results in the lowest weight Various methods of fabricating hollow steel and aluminum blades have been employed by different manufacturers Magnesium was used extensively for solid blades by the Germans in World War II but it has not proved very successful elsewhere Although magnesium has a weight advantage it has several important disadvantages such as poor fatigue erosion and corrosion qualities

Phenolic plastic reinforced with high strength fabric, has been used with moderately good success for low power applications Plastic has excellent vibration damping characteristics however the strength weight ratio is not so good as for some other materials

The hub parts of controllable pitch propellers are constructed with high strength alloy steel and aluminum

The structural problems encountered in the design of propellers are formidable because both the steady and vibratory forces are high Also the size and weight of the propeller must be kept to a minimum The chief steady forces are centrifugal thrust and torque Vibratory forces transmitted to the propeller hub by the engine produce alternating stresses in the hub structure which might result in fatigue failures Also the natural vibratory frequencies of the blades may be excited by various engine frequencies causing high vibratory blade stresses or by aerodynamic forces such as could occur when a blade tip passes near a part of the aircraft structure To guard against fatigue failures the hub parts are subjected to endurance tests at values of alternating and steady stresses substantially higher than those measured in service The alternating stresses existing in the blades during flight are measured by electronic strain gage equipment If the allowable limits determined from fatigue tests are exceeded the designer can either reduce the exciting forces by various means at his disposal or modify the blade so that the natural frequencies do not coincide with the exciting frequencies in the normal operating range of rotational speeds

Factors affecting efficiency The size shape and number of blades are usually determined from considerations of translating the engine power into thrust power in the most efficient manner ...

... best over all results diameter number of blades blade width blade thickness blade airfoil section planform pitch distribution and single or dual rotation The diameter is often limited by the available space or by the compressibility factor As the blade tips approach the speed of sound (or at about 9 Mach number) the drag of these tip sections increases very rapidly causing a substantial loss in efficiency Compressibility also limits the airspeed at which propell... can be ope...

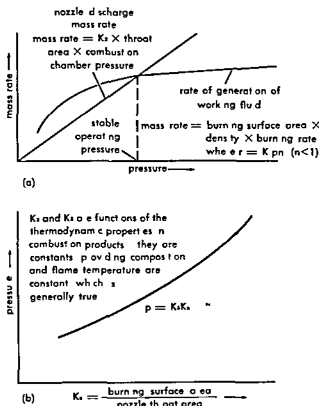


Fig 5 (a) Solid propellant reaches stable burning condition (b) Typical pressure characteristic for solid propellant

there seems to be no practical limit to the size of a solid propellant rocket [H W RI]

Bibliography J Humphries *Rockets and Guided Missiles* 1956 Princeton University Press *High Speed Aerodynamics and Jet Propulsion* vol 2 1956 vol 12 1959 G P Sutton *Rocket Propulsion Elements An Introduction to the Engineering of Rockets* 2d ed 1956 F A Warren *Rocket Propellants* 1959, M J Zucrow *Principles of Jet Propulsion and Gas Turbines* rev ed 1949

Propeller, air

A hub and multiblade device for changing rotational power of an aircraft engine into thrust power for the purpose of propelling an aircraft through the air An air propeller operates in a relatively thin medium compared to a marine propeller and is therefore characterized by a relatively large diameter and a fairly high rotational speed It is usually mounted directly on the engine drive shaft in front of or behind the engine housing

The propeller is an airfoil aerodynamically similar to a wing but rotated in the air directly by engine torque Motion of the propeller blade produces useful lift parallel to the axis of propeller rotation This lift produces the forward thrust that drives the aircraft through the air (*see* AIRFOIL)

Propeller types The simplest form of propeller is a two blade fixed pitch type used on relatively small low speed aircraft High speed aircraft, however require means for changing the blade angle during flight to obtain the highest efficiency for all flight conditions such as take off climb cruise and landing (Fig 1) Controllable-pitch propellers may be classified as two-position variable-pitch constant speed feathering and reversible-pitch The two position propeller is limited to two angular positions of the blade one for take off and climb and the other for cruise In the variable-pitch propeller the blade angle may be adjusted to any intermediate value between the low and high pitch limits in order to meet more precisely the requirements of all flight conditions The constant speed propeller is a variable pitch type having in addition a governor which automatically changes the pitch to maintain constant engine speed under varying flight conditions

The feathering propeller is a variable pitch or constant speed type capable of increasing the pitch beyond the normal high pitch value to the feathered position The latter is used on multi engine aircraft because the feathering feature provides a safety measure It is essential in the event of engine failure to stop the windmilling action of the propeller to reduce air drag by increasing the pitch to a point where this windmilling action and engine speed become zero This occurs when the blade angle is about 90° to the plane of rotation

The reversible-pitch propeller is a controllable or constant speed type having provisions for reducing the pitch to and beyond the zero value into the negative pitch range When the pitch is in the negative range the thrust is reversed which provides means for braking the aircraft during landing or means for maneuvering seaplanes on the water

Pitch control mechanisms The mechanisms employed in controllable pitch propellers for ro-

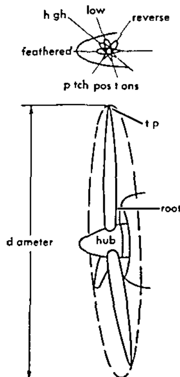


Fig 1 Principal parts of aircraft propeller and typical pitch positions

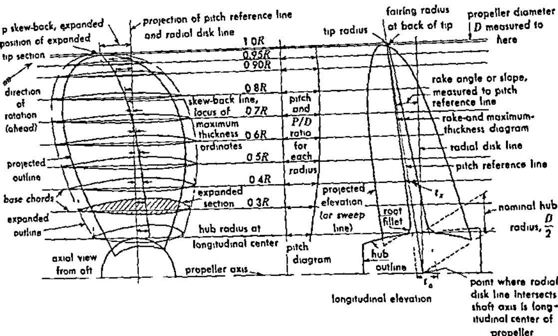


Fig 2 Drawing of screw propeller

Propeller geometry. Figure 2 shows a screw propeller drawing in which the usual orthogonal projections the blade surface as developed and expanded in a flat plane, and the cylindrical blade sections are shown The various technical terms used in discussion of screw propellers are indicated in the drawing The side of a propeller blade away from the hull is called the face or pressure side The opposite side is called the back or suction side

In its simplest shape the blade face is part of a helicoid surface and its axial advance per revolu tion defines the propeller's pitch which, in the case of a constant pitch propeller is constant over the radius R Many modern screw designs for reasons of better adaptation to the complex flow pattern behind the vessel s stern called the wake, employ a radially varying pitch (variable pitch propeller) The face pitch occurring at 0 7R is then taken to be representative for the mean face or nominal pitch (see Fig 2) The dimensionless ratio of the face pitch P to the propeller diameter D is the pitch ratio P/D It is an important char acteristic of the screw propeller

The difference between the pitch velocity nP (where n is revolutions/sec) and the mean axial speed of advance [illegible]
rou [illegible]
pre [illegible]
les [illegible]

$$S_R = (nP - V_A)/nP = 1 - V_A/nP$$

which is one of the fundamental quantities deter mining propeller action and efficiency

Propeller theory. The theory that led to the modern screw propeller developed mainly along two lines the momentum theory and the blade element theory In the former, the origin of thrust is explained by the change of momentum that takes place in the surrounding fluid (the slip stream) In the latter, the forces on each radial strip or blade element are determined and, by integration over the propeller radius, the total of their thrust and torque force components is calculated The momentum theory offers no indication as to the proper shape of the propeller, but yields useful general expressions for efficiency The blade element theory shows the effect of changes in shape of the propeller, but leads to untenable conclusions with regard to efficiency

[illegible] in the slip stream as being induced by vortex systems Developments of the circulation theory have called for highly complex design methods requiring the aid of electronic computers Simpler but useful approximate methods, which ensure good accuracy in design, are now also available for general application

Model tests. To aid in propeller design ex periments with systematically varied small scale propeller models were made This method intro duced by W Froude is still in general use and supplies the most direct and least time consuming approach to design and analysis The interac tion of hull and propeller can be studied in a towing tank through self propulsion tests with ship models See TOWING TANK

Number of propellers. In general a single screw arrangement yields a higher propulsive effi ciency than twin screws for fully loaded vessels, since the outboard shafting supports add considerably to the hull's resistance In ships such as

Propellers can be designed or selected for any particular application by several methods Families of propellers covering many of the various design parameters have been tested in wind tunnels and the data presented in dimensionless chart form Selection of the optimum propeller is simplified by the use of such charts Another method, the strip analysis method, is sometimes used for conditions where test data are not available In this method the blade is divided into a number of strips parallel to the airflow and the aerodynamic characteristics of these sections integrated to obtain the performance of the entire blade

A knowledge of the sources of propeller loss is a valuable tool in designing efficient propellers These losses are profile drag of blade sections, axial momentum loss in the slip stream, rotational momentum loss, and tip vortex loss The lowest sum of all the losses results in the highest efficiency Large diameters minimize axial and rotational momentum losses, which represent a large proportion of the total power at low airspeeds, but at high speeds these losses are of less importance Dual or counter rotation of propellers largely eliminates rotational losses, which become of considerable importance for cases where the power being absorbed is high in proportion to diameter Profile drag is always an important factor, which dictates the use of thin low drag sections which have good compressibility characteristics near the blade tips The blade width, or number of blades, is selected to provide the most efficient air foil section loadings

Under optimum conditions maximum efficiency of a propeller may reach 93% Most modern propellers operating on all but the smallest aircraft probably have efficiencies ranging between 85–90%

The propeller is responsible for most of the noise produced by a conventional propeller driven aircraft Propeller noise intensity which is greatest in the plane of the propeller, increases rapidly with tip speed particularly as the speed of sound is approached Noise is also a function of the power and number of blades The higher the power the more intense the noise but an increase in the number of blades reduces noise intensity (*see* AIRCRAFT NOISE) The noise decreases with increasing number of blades for constant power because the power absorption per blade decreases inversely as the number of blades and noise energy per blade decreases at a faster rate [D J B]

Bibliography N C Nelson *Airplane Propeller Principles*, 1944, F E Weick, *Aircraft Propeller Design* 1930, U S Defense Dept, *Aircraft Propeller Handbook*, ANC-9, 1956

Propeller, marine

A component of a ship propulsion power plant which converts engine torque force into propulsive force or thrust, thus overcoming the hull resistance of a moving ship by creating a sternward accelerated column of water Historically preceded in steamship propulsion by the jet propeller (1782) and the paddle wheel (1801), the screw propeller (1804) gradually superseded the earlier propeller forms and since 1860 it has been the only propeller type used in ocean traffic mainly because of the evolution of the marine engine towards higher rotative speed

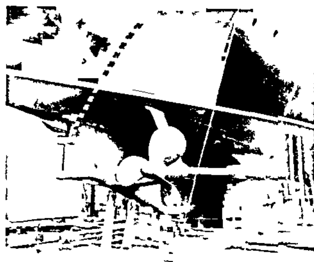


Fig 1 Stern view of twin screw SS *Florida* showing location of three bladed propeller and its shaft The other propeller is hidden from view by the rudder (From D Arnott, ed, *Design and Construction of Steel Merchant Ships, Society of Naval Architects and Marine Engineers, 1955*)

The advantages of a screw propeller include light weight, flexibility of application good efficiency at high rotative speed and relative insensitivity to ship motion The fundamental theory of screw propellers is applicable to all forms of marine propellers In its present form a screw propeller consists of a streamlined hub attached outboard to a rotating engine shaft on which are mounted two to six blades The blades are either solid with the hub detachable or movable The screw propeller has the characteristic motion of a screw, it revolves about the axis along which it advances The screw blades are roughly elliptical in outline One or more screw propellers are usually fitted as low as possible at the ship s stern to act as thrust producing devices (Fig 1) The low position of the propellers affords good protection and sufficient immersion during pitching movements of the ship The screw diameter should not be larger than 0 7 of the loaded draft in seagoing single screw vessels *See* SHIP DESIGN

In towboats developing high thrust at low transport speed the towing force can be increased up to 40% by shrouding the screw in a coaxially converging streamlined nozzle (Kort nozzle) In combination with a tunneled stern this device finds wide application in powerful river towboats of modern design If hinged about a vertical axis the nozzle can also serve as a powerful steering device (nozzle rudder)

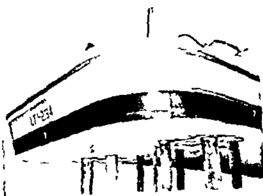


Fig 3 Twin cycloidal propellers on U S Army tow boat The twin units are rated at 1100 hp each Blade length is 4½ ft (Pacific Car and Foundry Company)

ing unusual flexibility and maneuverability at low

FRY WAKE (SEAGOING VESSELS) [L TR]

Bibliography M K Eckhardt and W B Morgan, A propeller design method *Soc Naval Architects Marine Engrs Trans* 63 325–374 1955 H W Lerbs Moderately loaded propellers with a finite number of blades and an arbitrary distribution of circulation *Soc Naval Architects Marine Engrs, Trans* vol 60 1952 H E Rossell and L B Chapman (eds) *Principles of Naval Architecture,* vol 2 1939 H E Saunders *Hydrodynamics in Ship Design* vols 1 2 1957 M St Denis and J P Craven Recent contributions under the Bureau of Ships fundamental hydromechanics research program *J Ship Research* 2(2) 1 36 1958 L Troost, Open water test series with modern propeller forms *Trans North East Coast Inst Engrs and Ship builders* 67(3) 89–103 1951

Properdin

A serum protein aiding the body defenses against microorganisms It was described by Dr Louis Pillemer in 1954 His material was a euglobulin with a molecular weight greater than 10^6 and an isoelectric point between pH 5 5 and 5 8 The properdin system consists of properdin magnesium ion (Mg) and all four components of complement C1 C2 C3 C4 Properdin is bactericidal to many although not all gram negative bacteria and it reacts with a variety of tissue and microbial polysaccharides including yeast zymosan with resulting inactivation of the complement Pillemer believed the third component of complement (C3) was largely affected by the yeast zymosan Dr R A Nelson Jr has advanced evidence that all four components are about equally reduced Pillemer defined the properdin *unit* as that amount which with excess zymosan would inactivate 120 ± 30 units of C3 in 1 hour at 37°C This corresponds to 0.5 μg properdin nitrogen Normal human serum contains 4-8 units per milliliter or not more than 0 02% of its total protein Properdin activity has been found even in the serum of germfree rats and in human agamma globulinemia It varies little in a variety of disease states although transient changes in the serum activity have been noted in pneumococcal pneumonia and in meningococcemia and also following total body irradiation

Pillemer believed that properdin was distinct from any conventional antibody Dr Nelson and his coworkers have however, suggested that properdin consists of one or more antibodies of broad specificity that cross react with cell constituents of certain microorganisms and that this combination is strengthened by further reaction with complement See AGAMMA GLOBULINEMIA, ANTIBODY, COMPLEMENT (SERUM), GERMFREE VERTEBRATE, PNEUMOCOCCUS, PROTEIN, RADIATION INJURY (BIOLOGY), SERUM [H P T]

Bibliography O S Whitelock (ed) *Conference on Natural Resistance to Infections,* Ann N Y Acad Sci, 66 233–414 1956

Propionibacteriaceae

A family of anaerobic bacteria of the order Eubacteriales This family contains relatively anaerobic, nonmotile nonsporulating gram positive predominantly rod shaped bacteria that ferment glucose and certain other substrates Some of the Propionibacteria are involved in the production of Swiss cheese and others in the production of vitamin B_{12} Three genera are included in the family, *Propionibacterium, Butyribacterium* and *Zymobacterium*

Propionibacterium This is a genus of mostly nonpathogenic bacteria that form propionic acid as a major product of their energy metabolism About 11 species are included in this genus A few propionic acid forming bacteria such as *Clostridium propionicum* and *Veillonella alcalescens* are placed in other genera *V alcalescens* is also known as *Veillonella gazogenes* and *Micrococcus lactilyticus* The cells of *Propionibacterium* species are predominantly short rods, but under some conditions they become almost spherical resembling streptococci and under other conditions particularly when growing in the presence of oxygen they develop as irregular club shaped or branched rods Propionibacteria grow best in the absence of oxygen but they tolerate low oxygen concentrations Relatively complex media containing a variety of amino acids and growth factors in addition to an energy source are most favorable for growing these bacteria As energy sources most species can use sugars or ... mannitc ... acids s ... a sugar or polyalcohol is used as substrate acid is formed, this must be neutralized by the addition of a suitable buffer to obtain abundant growth of *Propionibacterium*

The fermentation of lactate results in the formation of mainly propionate acetate and carbon dioxide according to the equation

tankers however the twin screw arrangement may compare favorably in the ballast condition completely offsetting the loss in the loaded condition

Turbine horsepowers exceeding 50 000 can be successfully transmitted by a single screw in large and fast ocean going ships Greater power output necessitates the use of multiple (mostly twin or quadruple) screw arrangements In some cases as in powerful shallow draft towboats the maximum available propeller diameter limits the efficient conversion of engine power per shaft In other cases preference for a certain number of diesel engine units to produce the propulsive power needed determines the number of propellers

Relative to single screw propulsion multiple screw arrangements in general give rise to greater machinery cost and weight with conventional machinery However considerations of operational advantages or of safety with regard to engine break down (especially in the case of large passenger vessels) may favor multiple screw arrangements

Number of blades Both the maximum open water efficiency and the optimum blade diameter for highest propulsive efficiency decrease slightly with increasing number of blades In practice the number of blades is from two to six and three or four are most common Two bladed screws are used in fast racing craft and in sailing vessels with auxiliary engine power where the propeller can be locked in upright position behind the stern Three bladed propellers are often used in multiple screw arrangements but design considerations may favor four bladed propellers of smaller diameter and involving smaller and lower resistance appendages Center screws unless high rotative speed prevails are mostly four bladed

In large and full ships with greatly variable wake five bladed screws are finding some preference because of greatly reduced thrust and torque variations Consideration of the natural frequencies of vibrations of the hull and the propulsive plant and their major harmonics in the operating speed range is the final controlling factor Excitation of resonances at blade frequency should be avoided Generous clearances between the propeller sweep line and the hull and its appendages are essential for reducing these excitation forces For additional information on ship vibrations *see* SHIP PROPULSION

Propeller cavitation The formation of vapor filled bubbles called cavitation causes noise vibrations and often rapid erosion of the propeller material especially in fast high powered vessels As long as the rotational and translational speeds of the propeller are not too high the onset of cavitation can be delayed by clever design of blade sections The application of circulation theory in design is of great help but it is strongly advisable to have a propeller model tested in a water tunnel under simulated operating conditions and at the proper cavitation index *See* WATER TUNNEL

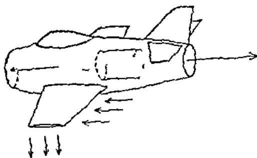
At exceedingly high ship speeds (of the order of 50 knots or greater) cavitation can no longer be avoided Theoretical and experimental developments however have indicated that reasonably good propulsive efficiency can still be obtained by propellers designed to operate with the backs of the blades in a fully developed cavity (supercavitating propellers) Since collapse of the cavitation bubbles no longer takes place on the propeller blades under these circumstances there is no erosion danger The blade sections of such propellers are wedge-shaped with a sharp leading edge and a blunt trailing edge For additional information on supercavitating propellers and cavitation phenomena *see* CAVITATION

In cavitation endangered propellers it may be advisable to use materials of greater corrosion resistance than the usual manganese brass There are now zinc free nickel aluminum bronzes available that are in addition lighter and stronger hence permitting the use of finer blade sections Certain stainless steel alloys are also finding increasing application for heavy duty propellers

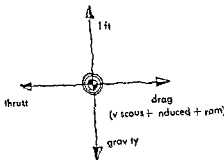
Propeller vibrations Propellers are usually mounted at the end of long shafts Such systems may be driven into several vibration modes by periodic flow and engine torque impulses These modes can be torsional axial and flexural and can occur singly or in combination Stern lines conducive to even flow into the propeller and generous propeller clearances are desirable features for good performance Certain blade vibrations produce objectionable grinding or piercing noises (singing propellers) In most cases singing of propellers can be prevented or eliminated by the application of so called chisel edges to provide fixed separation points along the trailing edges of the screw blades

Controllable pitch propellers For ships which normally operate at widely diverging speeds and propeller loadings (towboats rescue vessels trawlers ferry boats) the application of controllable pitch (movable blade) propellers permits the use of full engine power at rated rpm under all operational conditions insuring maximum thrust production utmost flexibility and maneuverability Since these propellers are also reversible they permit the use of nonreversible machinery (gas turbines) The hydraulic or electric servomotor for adjusting the pitch of the blades requires for its operation a hollow tailshaft The propeller pitch can be directly controlled from the ship s bridge In each individual case the operational advantages of the controllable pitch screw must be weighed against the disadvantages of more complex construction and higher manufacturing cost

Cycloidal propellers A type of propeller wheel carrying near its periphery a number of spadelike vertical blades which perform a rotary movement about their vertical axes (Fig 3) Depending on the disposition of the blades the propeller s thrust can be directed ahead sideways or astern provid



Air momentum



reaction forces

Air momentum and reaction forces

momentum resulting in upward reaction forces called lift on the airplane The air masses deflected downward are also given some forward momentum in the process accompanied by rearward reaction forces against the airplane called induced drag The airplane propulsion system takes in air accelerating it to approximately the forward speed of the airplane The rearward reaction forces resulting from this momentum change are called ram drag The air taken into the propulsion system is then discharged as a rearward jet with a relatively great momentum resulting in forward force on the airplane called propulsive thrust

For the situation described momentum is conserved in directions parallel to the line of flight Consequently all the forces acting on the constant speed airplane offset one another so that the net external force is zero

Thrust The capability of a propulsion system is measured in terms of the net thrust it delivers to the vehicle In the case of an airplane powered by a thermal jet system such as a ram jet or turbojet the net thrust is *the difference between the propulsive thrust* and the ram drag *See* JET PROPULSION

From a point of reference on the moving airplane the net thrust is

$$F = \frac{(w_a + w_f)V_r}{g} - \frac{w_a V_p}{g} \qquad (2)$$

where F = net thrust lb
V_r = relative velocity of air leaving the propulsion system ft/sec
V_p = forward airplane velocity ft/sec
w_a = weight flow of air lb/sec
w_f = weight flow of fuel lb/sec
g = acceleration of gravity ft/sec²

The first term is a measure of the total propulsion system thrust and the second term is a measure of the system ram drag

In a rocket engine no air is taken into the propulsion system consequently there is no ram drag The net thrust of such a system is therefore the first term of Eq (2) with w_a replaced by w_o, the weight flow of oxidizer lb/sec

Power and efficiency The over all suitability of a propulsion system is related to two factors (1) how efficiently it converts energy supplied to the system into propulsion power and (2) how efficiently this power is used in driving the vehicle

The first factor thermal efficiency η_{th} of the propulsion system is defined as

$$\eta_{th} = \frac{\text{propulsion power developed}}{\text{rate at which energy is supplied}}$$

The second factor propulsive efficiency η_p is defined as

$$\eta_p = \frac{\text{useful thrust power}}{\text{propulsion power}}$$

The over all efficiency η of the propulsion system is

$$\eta = \eta_{th}\eta_p \qquad (3)$$

Useful thrust power is the rate of doing useful work on the vehicle and is the product of net thrust and vehicle velocity

$$P_T = FV_p = \frac{w}{g} V_r V_p \qquad (4)$$

where P_T = thrust power ft lb/sec
F = net thrust lb
$w = w_a + w_f$ for a thermal jet or
$w = w_f + w_o$ for a rocket

The propulsion power developed consists of the useful thrust power plus the power lost by dissipation or ejection in a manner that does not contribute to useful thrust power Thus

$$P = P_T + P_L \qquad (5)$$

where P is propulsion power and P_L is power lost

In the ideal case P_L consists only of the rate of kinetic energy dissipated by the working fluid due to its absolute velocity rearward after being discharged from the propulsion system In this case

$$P_L = \frac{w}{2g}(V_r - V_p)^2 \qquad (6)$$

Stated in terms of working fluid flow rate jet velocity and vehicle velocity the propulsion power is



$$3CH_3CHOHCOOH \rightarrow$$

Lactic acid

$$2CH_3CH_2COOH + CH_3COOH + CO_2 + H_2O$$

Propionic acid Acetic acid Carbon dioxide Water

In addition small amounts of succinate a dicarboxylic acid and propyl alcohol are commonly formed When glycerol is used as a substrate the main product is propionic acid with lesser and variable amounts of succinic acid



$$CH_2OHCHOHCH_2OH \rightarrow CH_3CH_2COOH + H_2O$$

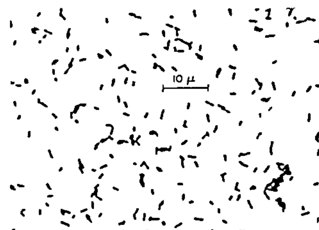
Glycerol Propionic acid Water

The yield of succinic acid in the fermentation of glycerol is greatly dependent upon the amount of carbon dioxide in the medium increasing with the carbon dioxide concentration Carbon dioxide is consumed by these nonphotosynthetic bacteria 1 mole of carbon dioxide being used in the formation of 1 mole of succinic acid The propionic acid bacteria were among the first heterotrophic organisms shown to consume carbon dioxide in substantial amounts

Propionic acid bacteria have been isolated chiefly from cheese milk and other dairy products They are also found in smaller numbers in soil decomposing plant materials such as silage and in the digestive tract of ruminants Propionic acid bacteria are particularly abundant in Swiss (Emmentaler) cheese During preparation of this type of cheese the lactose is first fermented by lactic acid bacteria and then the lactate is slowly fermentated by propionic acid bacteria The carbon dioxide gas produced by the latter organisms is responsible for the formation of the characteristic holes in the cheese and the characteristic sharp flavor is partially attributable to the presence of propionic acid *See* CHEESE

Some species of *Propionibacterium* form relatively large amounts of vitamin B_{12} and are used for its commercial production *See* VITAMIN B_{12}

Butyribacterium This is a genus containing bacteria that form butyric acid acetic acid and carbon dioxide as major fermentation products The only species that has been fully described is *B rettgeri* This organism was isolated from intestinal contents of a white rat It requires a complex medium for growth and is able to ferment glucose maltose and lactate The ability to form butyric acid from lactate distinguishes organisms of the genus *Butyribacterium* from *Lactobacillus Propionibacterium* or *Zymobacterium* species



Butyribacterium rettgeri

Zymobacterium This is a genus containing bacteria that form predominantly ethyl alcohol and carbon dioxide from glucose Only one species *Z oroticum* has been described This organism is also able to ferment orotic acid a compound involved in the biosynthesis of the pyrimidines of nucleic acids Some of the early steps in pyrimidine biosynthesis were discovered during studies of the decomposition of orotic acid by cell free extracts of *Z oroticum* *See* BACTERIA TAXONOMY OF, EUBACTERIALES [H A B]

Bibliography K V Thimann *The Life of Bacteria* 1955

Propulsion

The process of causing a body to move by exerting force against it Propulsion is based on the reaction principle (Newton s third law) which states that for every action there is an equal and opposite reaction

Vehicles moving in fluid media are usually propelled by devices or systems which operate on the media For example a ship is driven forward by reaction to the force exerted against the water by its propeller Similarly an airplane may be propelled by reaction to force exerted against air passing through its turbojet propulsion system A rocket propulsion system however does not depend upon surrounding media its working fluid is the hot gas resulting from the burning of fuel and oxidizer which it carries along

Momentum theory The propulsion system of a ship airplane or rocket imparts momentum to the fluid upon which it operates The magnitude of force applied to the fluid and consequently the driving reaction force applied to the vehicle are related to the momentum change imparted to the fluid according to Newton s second law The rate of change of momentum is proportional to the force applied to the body and takes place in the direction of the line of action of the force Thus

$$F = dM/dt \quad (1)$$

where F is force lb, M is momentum (mass times velocity) slug sec and t is time sec Consider the case of an airplane in level flight at constant speed as illustrated The air is disturbed in several ways by its passage Some masses of air are given forward momentum through viscous adherence to external surfaces of the airplane resulting in rearward forces on the airplane called viscous drag Some masses of air pass across the wings and other surfaces of the airplane and are given downward

lial outgrowth (Patellidae) or a pulmonary chamber (Helicinidae and Hydrocenidae) The visceral loop is in a figure 8 (streptoneurous) An operculum is usually present there is generally only one pair of tentacles and the sexes are usually separate The shells produced have evolved many shapes from the caplike limpets in the genera *Patella* and *Acmaea* without a coil the slipper limpets in the genus *Crepidula* with only a slight indication of a coil to the spiral shells of many whorls such as the auger snails in the genus *Terebra* The embryonic shell called the protoconch develops during the larval veliger stage and is composed mainly of periostracum This is followed by the later whorls of lime which usually have the periostracum continued as an outer layer

The more primitive species are plant feeders feeding mainly on algae and other marine plants The more advanced members of this subclass are scavengers or predators feeding upon other mollusks marine worms sea urchins and even fish A limited number of the predatory species by means of their radula can bore through the shells making a small hole through which they introduce the proboscis and feed upon the soft parts Many of these predatory groups especially in the genera *Urosalpinx Thais Eupleura* and *Murex* are of considerable economic importance These predators are called oyster drills because they destroy many young oysters their main source of food.

Certain of the primitive forms have nacreous ...

... this subclass have highly lustrous shells such as the Cypraeidae Olividae and Volutidae In these families the mantle can be extended over the outer surface of the shell laying a laminated layer of lime on the outside as well as on the inside In cross section ...
laminated ar...
Members of ...
colored and ...
PIDOBRANCHIA ... NEOGASTROPODA PECTINIBRANCHIA

[W J C]

Prospecting

The search for mineral deposits that can be worked at a profit A prospect is an occurrence of mineral ...

... containing metallic elements such as copper lead zinc or iron nonmetallic materials such as asbestos clay phosphates or sulfur and mineral fuels such as coal or petroleum Deposits worked for their aggregate of materials such as sand gravel or dimension stone are usually considered deposits of the rock itself For a discussion of general geological and geophysical methods of search for petroleum deposits see PROSPECTING ...
has ... common

metals and nonmetals by traditional methods so that in most countries the more obvious deposits have been found Hope for future discovery therefore, lies mainly in the buried deposits which at best may give only subtle indication of their presence at the surface

Not only are the older field and library methods being systematically improved but the need to prospect for buried deposits has stimulated the use of indirect methods In these geophysical geochemical and botanical evidences of subsurface conditions supplement surface evidence Because of increased complexity and capital requirements, an increasing proportion of prospecting is done by groups of specialists working for mining companies Such a group may consist of a geologist and an engineer, and assistants as required for traversing and sampling Specialists in photogeology (see AERIAL PHOTOGRAPH) and in geophysical and geochemical prospecting may be called in as needed Initial prospecting is followed where warranted by trenching or drilling and further sampling For principles of mine evaluation and field sampling see MINING OPERATING FACILITIES

The individual prospector independent or employed by mining companies remains important, but probably less so than in the past More and more he will need a working knowledge of geology of the ...

... principles of photogeology, be able to make mineralogical and geochemical tests in the field understand the use of the simpler geophysical instruments and make use of vegetation rock staining and float as guides to mineralization

Outline of prospecting methods Mineral prospecting normally proceeds from the general to the specific from ...

... including a study of available maps and reports prospecting may often be concentrated on smaller areas immediately Prospecting methods may be subdivided into direct and indirect methods Direct methods include geologic and photogeologic mapping, study of guides to ore, and field examination of the surface supplemented by panning, trenching pitting drilling or sampling Indirect methods are of two kinds (1) geophysical methods which include magnetic electromagnetic and radioactivity surveys both from the air and on the surface electrical resistivity self potential gravimetric and seismic surveys on the surface, and electrical self potential radioactivity, and temperature surveys in boreholes (see GEOPHYSICAL EXPLORATION) and (2) geochemical and botanical surveys

Where there is surface evidence of minerals examination and sampling may be all that is necessary to determine if further exploration is warranted In little-known regions or where the deposits are deeply oxidized leached or buried prospecting

$$P = \frac{w}{g} V_r V_p + \frac{w}{2g}(V_r - V_p)^2 \qquad (7)$$

From these definitions ideal propulsive efficiency η_P may be expressed as follows

$$\eta_P = \frac{P_T}{P} = \frac{w_a - w/\beta}{(w_a - w/\beta) + \frac{W}{2}(1 - 1/\beta)^2} \qquad (8)$$

where $\beta = V_p/V$ the ratio of vehicle and working fluid velocities For turbojet or propeller systems $w_f \approx 0$ hence

$$\eta_P = \frac{2\beta}{1+\beta} \qquad (9a)$$

For rocket engines $w_a = 0$ hence

$$\eta_P = \frac{2\beta}{1+\beta^2} \qquad (9b)$$

The closer the vehicle velocity approaches the relative velocity of the ejected working fluid the greater is the propulsive efficiency so that $\eta_P \rightarrow 1$ when $\beta \rightarrow 1$ However in the case of a propeller or thermal jet system where air is the working fluid the net thrust is proportional to the difference in vehicle and relative jet velocities and decreases as $\beta \rightarrow 1$ Because there must always be a positive net thrust to overcome viscous and induced drag such a system to be practical must operate at less than perfect propulsive efficiency $\eta_P < 1$ because $\beta > 1$

In the case of a rocket however the net thrust of the system depends entirely on the relative leaving velocity V of the gases and such a system may attain unity propulsive efficiency when $\beta = 1$ in which case all the propulsion power is converted into thrust power

Specific fuel consumption Frequently in considering the suitability of a thermal jet propulsion system the specific fuel consumption (SFC) is referred to This is the fuel consumed per hour per unit net thrust delivered For a thermal jet it is

$$\mathrm{SFC} = 3600g\,\frac{w_f}{w_a}\left(\frac{1}{V_r - V_p}\right) \qquad (10)$$

where SFC is pounds of fuel per pound of thrust per hour

Specific impulse Another factor to be considered in determining the suitability of a propulsion system is its specific impulse the thrust delivered per unit weight flow of working fluid in a given system

For a turbojet or other thermal jet where $w_f \approx 0$ it is convenient to refer to air specific impulse

$$I_{a\,r\,sp} = \frac{F}{w_a} = \frac{V_p}{g}\left(\frac{1}{\beta} - 1\right) \qquad (11a)$$

The size and weight of a turbojet propulsion system are determined to a considerable degree by the volume of air flow it must handle to produce the required thrust Hence a system having high relative jet velocity is to be desired for the most compact arrangement However such a system has poorer propulsive efficiency if its jet velocity is large compared to the airplane velocity The choice of an optimum system usually involves a compromise between these opposing factors determined largely by speed altitude and duration of the intended airplane mission

For a rocket propulsion system the specific impulse is

$$I_{sp} = \frac{F}{w_f + w_o} = \frac{V_r}{g} \qquad (11b)$$

This measures the thrust delivered per unit weight flow of propellants (fuel plus oxidizer) and is not related to the velocity of the vehicle No compromise with propulsive efficiency is necessary thus for the total thrust required a system giving the greatest possible specific impulse is to be preferred *See* AIRPLANE ROCKET ENGINE SHIP PROPULSION TURBINE PROPULSION [D C]

Bibliography C W Smith *Aircraft Gas Turbines* 1956 M J Zucrow *Aircraft and Missile Propulsion* vol 1 1958

Propylene

A colorless gas C_3H_6 boiling point −47 7°C melting point −185°C All processes for thermal or catalytic cracking of hydrocarbons yield propylene A typical fraction of a refinery stream containing 2 and 3 carbon molecules is 10-30% propylene Preparation outside of normal refinery operations is usually by catalytic dehydrogenation of propane In those refineries having integrated isopropyl alcohol or polymer units for making polymer gasoline the propylene is utilized in the dilute concentration found in the cracked gas streams When highly concentrated streams are required the propylene is recovered in the same manner as ethylene Major uses for propylene are in the production of isopropyl alcohol from which acetone is obtained tripropylene for the alkylation of phenol to produce alkyl phenols from which are derived nonionic detergents and lubricating oil additives tetra propylene for the alkylation of benzene to produce alkyl benzenes from which are derived alkyl aryl sulfonate detergents propylene chlorohydrin for producing propylene oxide from which are obtained detergents hydraulic fluids and lubricants acrolein allyl alcohol allyl chloride epichlorohydrin synthetic glycerin butyraldehyde *n* butyl alcohol and isobutyl alcohol *See* ALKENE ETHYLENE POLYOLEFIN RESINS [C A C]

Prosobranchia

The largest subclass in the class Gastropoda containing most marine snails the limpets and whelks a limited number of nonpulmonate land snails and nearly all of the nonpulmonate fresh water snails Three orders are commonly recognized the Aspidobranchia Pectinibranchia and Neogastropoda

Respiration is usually by means of ctenidia or gills located in front of the heart When gills are absent respiration is carried on by means of a pal

may bear no relation to the present drainage In areas where rock outcrops are scarce assemblages and shapes of minerals in stream channels are indicative of the nature of the parent rocks and of the possible occurrence of economic minerals

Float (detatched fragments of mineralized rock or vein material) indicates a bedrock source at some higher point which may be found by systematic search Finer material in residual overburden or in alluvium may be traced to its source by panning Most ore minerals are destroyed rapidly by weathering so their source is likely to be close if they are found in float or by panning The more resistant materials such as vein quartz, gold cassiterite chromite tantalite rutile and zircon may be carried far from their sources Ore minerals are genetically associated with minerals that are of no economic importance but may be relatively abundant or resistant and thus indicate the presence of the valuable minerals As examples chromite and titaniferous magnetite are associated with metals of the platinum group pyrite pyrrhotite chalcopyrite and vein quartz with gold and ilmenite magnetite chromite and pyrope (deep red garnet often a gem stone) with diamonds Similarly fine gold in residual soil may indicate a bed rock source of copper farther uphill as copper sulfides decompose readily leaving free gold in the residue Float tracing and panning work best in unglaciated country

Testing and sampling At the point where the mineral indications go beneath the surface trenches or test pits must be dug or bulldozed Ground sluicing or hydraulicking may be used to advantage where water is abundant, slopes are moderately steep and stream pollution is unobjectionable Drilling is sometimes necessary in prospecting areas covered by swamp water or rock or where deep overburden makes trenching or pitting unduly expensive Drilling is useful in searching for bedded deposits buried placers, and extensions of known ore bodies and for locating faults or faulted segments of ore bodies Drilling is not well suited to prospecting for narrow veins friable ores or small irregular ore bodies See BORING AND DRILLING MINERAL

Prospects are sampled to determine their composition Methods for representative sampling depend on the size character and accessibility of the project Exposed prospects are mostly sampled by chip grab or channel samples Buried deposits are commonly sampled by trenching or pitting or by the drilling methods Placers are sampled by pan or rocker

INDIRECT METHODS

In a large proportion of indirect prospecting geophysical methods are used involving the measurement of physical quantities associated with buried mineral deposits and geological structures The measurements are interpreted in terms of geology Plausible assumptions are made about the subsurface and the physical effects of assumed structures or deposits are computed or estimated and compared with the geophysical measurements The assumptions are modified until there is reasonable agreement between the computations and observed data

Selection of a suitable geophysical method is facilitated by information on the geologic habits and environment of the deposits sought and on their physical properties compared with the surrounding rocks Magnetic methods and the various electrical methods are commonly used in prospecting for metallic minerals because these minerals usually have magnetic and electrical properties that contrast with those of the surrounding rocks. Seismic and gravitational methods have been less used because measurable contrasts in elasticity and density are less common but they have been indirectly useful in locating buried structures Little information on depth is theoretically obtainable with magnetic, gravitational self potential and thermal methods where physical effects are produced by the bodies or structures themselves. Information on depth can usually be obtained with resistivity or seismic methods where effects are produced by transmitting electrical or seismic energy through the ground

Geophysics has been used less in prospecting for metallic and nonmetallic minerals than for petroleum because of the small size of most ore bodies compared with oil structures and the great variety in their shape physical properties and geologic environment

Geophysics has nevertheless been used increasingly in mineral prospecting because of the increased emphasis on the search for buried deposits The development of air borne magnetic electromagnetic and radioactivity methods make it possible to cover large and otherwise inaccessible areas at comparatively low cost All three types of surveys may be made simultaneously from the same aircraft at little more than the cost of a survey by a single method

Magnetic methods Air borne and ground magnetic surveys have been used extensively to prospect for magnetic deposits and for nonmagnetic deposits which either contain magnetic gangue minerals or are associated with structures or bodies that have magnetic expression Contact metamorphic and replacement deposits commonly contain magnetic gangue minerals and may thus be indicated by magnetic surveys Faults dikes contacts of igneous intrusions lava beds and magnetite-bearing sedimentary and metamorphosed rocks wh[illegible] many c[illegible] anomali[illegible] ated wi[illegible] [illegible] es of alteration in igneous rocks

In aeromagnetic surveys flying elevation and spacing of flight lines are governed by terrain and by the geologic habits and associations of the deposits sought Flights may vary from one eighth mile apart when prospecting for discontinuous magnetic deposits to one mile or more in tracing a

must be based on geologic inference and indirect methods and operations become more complex

DIRECT METHODS

A map of some sort is required to prospect large areas systematically Government geologic and topographic maps if available are suitable bases for plotting mineral occurrence or guides to minerals Aerial photographs are excellent for such purposes though ground control is necessary to produce maps from them Photogeologic studies combined with ground checking may indicate such guides as soil staining alteration or structures in deeply weathered areas that might otherwise be missed Recent work has shown that in some areas aerial observation and color photographs may be useful

Ore guides These are mostly associations of geologic and other regional factors Recognition of ore guides is highly important though not all guides are valid everywhere The ore guides that are significant depend on the characteristic rock associations and distribution of the ore in any given region or locality For example mineral stream gravel or placer deposits generally indicate ore in vein or lode in the country rock of the back land

Among regional ore guides of general rather than specific application are large igneous intrusions with which ore is known to be associated Examples are the Sierra Nevada batholith of California and the Coast Range batholith of British Columbia Copper tungsten tin and molybdenum are characteristically associated with granitic rocks nickel with mafic rocks especially norite chromium nickel and platinum with ultrabasic igneous rocks beryllium with pegmatites and uranium vanadium and selenium with terrestrial and shallow water beds of sandstone and shale

Faulting and rock weathering offer clues to mineralization Major zones of faulting may be valid regional guides although ore deposits are more commonly found in subsidiary faults related to major fault zones The Mother Lode of California a mineralized fault and shear zone extending 120 miles along the western side of the Sierra Nevada batholith is the best known example in the United States Bauxite manganese barite and lateritic iron ores are found in deeply weathered rocks The porphyry copper deposits of the southwestern United States and western South America occur where arid climate and deep water level have favored oxidation of the original material and preservation of zones of enrichment Deep weathering and subsequent active erosion favor concentration of gold platinum ilmenite zircon monazite and some rare earths in placers

Ore deposits also are characteristically associated with so-called metallogenetic epochs when conditions favored introduction concentration and deposition of minerals (*see* ORE AND MINERAL DEPOSITS) Ore minerals deposited during Precambrian metallogenetic epochs are important in eastern North America especially in the Canadian Shield in the shields of Brazil Scandinavia Finland and southern Siberia and in ancient rocks in Africa and other parts of the world Minerals deposited during Cretaceous and Tertiary epochs are important in western North America including most of Mexico and Alaska and in western South America

Local ore guides include rock alteration and channelways such as fractures faults contacts between dissimilar rocks and breccia pipes Ore shoots are likely to occur where mineralizing solutions encounter easily replaced rock such as limestone where veins cross contacts at intersections of veins and faults at rolls in faults and along veins wherever there is a change in physical or chemical environment

Any departure from normal structure topography rock color or vegetation should be investigated Old mine workings dumps prospect pits and burrows of animals may yield information on mineralization and on the size and trend of deposits

The gangue minerals in veins almost invariably extend far beyond the limits of ore and thus constitute important guides Alteration halos in materials adjoining veins are also important though alteration may be so widespread that it serves merely as a regional guide Pyrite white mica and clay are common and easily recognized alteration minerals but microscopic study may be necessary before others can be identified Rock staining may result from weathering of sulfides of iron manganese copper cobalt and nickel and the oxides of uranium and vanadium Each leaves a characteristic color red brown to pale yellow from iron black from manganese green and blue from copper pink to red from cobalt apple green from nickel bright yellow orange or green from uranium and yellow from vanadium

Gossan (mainly hydrated iron oxides) may be the surface residue of an ore deposit from which sulfides of copper lead or zinc have been leached Base metal sulfides commonly leave cellular pseudomorphs when leached from the primary ore so that the structure of the gossan may tell much about the original mineral content Native gold silver copper and bismuth and locally oxides of copper lead zinc and manganese tend to remain in gossan The gold content of gossan may in fact be higher than the underlying primary deposits

The topography of a region may furnish useful clues though no general statement can be made concerning the relation between mineral deposits and topography Outcrops of ore gossan or even sparsely mineralized gangue are of course directly useful

Placers are formed by weathering erosion and concentration in channels of resistant heavy metals and minerals such as gold platinum cassiterite rutile and diamonds and other gem minerals An understanding of the processes involved will therefore aid in recognizing ancient channels which

able contrast in conductivity The applications are diverse as each type of deposit presents its own problem

A variety of instruments has been u ed in elec tromagnet c prospecting, but all measure anoma l es in the electromagnetic field set up by induced ground currents rather than in the potential dis tribution as in resi tivity methods The frequency of the energizing current *should be 250–1500* cy cles per second Too low frequencies reduce the strength of the induced field but too high frequen cies lack depth penetration

A recent development uses natural random alter nat ng magnetic fields of audio and sub audio fre quency to locate subsurface electrical *conductors* such as sulfide ore bodies Distortions of the fields caused by variations in conductivity are measured at *several frequencies* by use of search coil detectors *The new method may be used for air borne or* ground surveys Although a high degree of amplifi *cation* is required theoretically greater lateral and *vertical* range is possible than with most electro magnetic methods

Air borne and ground electromagnetic surveys *ha e been useful* mainly in areas where unaltered sulfide deposits are fairly close to the surface as in the *Canadian* Shield Other conductors such as *fault* gouge graphite and some water bearing for mat ons may *likewise be indicated* by electromag net c surveys Air borne electromagnetic surveys

to flat lying deposits or beds may often be determined fairly accurately using specialized mathematical methods of interpretation The induced potential method is a modified electrical method that shows prom e in prospecting for metallic ores.

Electrical and ele

ber

ma

two or more boreholes have als been successful in locating conducting ore bod es The pr nciples are identical with those gov ern ng s rface surveys

A spontaneous polarization or so-called

ected to two nonpolarizable *electrodes* The accuracy of measurements is re duced by variations in soil type and soil solutions roots of trees and topography which may set up potent als larger *than those* associated with some re bod es *Local* spurious potentials can be elimi nated largely by *repeating the* observations with the electrodes in *different* positions

Self potent al surveys are useless where the de po ts are deeper than 100 ft or where they are be low the zone of *oxidation In the latter* case good results have been obta ned when the water table is lowered during long dry periods and the sulfide *minerals* are subject to renewed oxidation

Gravity, seismic and thermal methods Most mineral deposits are too small or have *insufficient density* contrast to permit use of gravity methods A further disadvantage is the requirement to cor

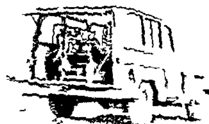
domes with their associated sulfur deposits on the Gulf Coast They have also been used to a minor ex tent in prospecting for relatively dense deposits such as magnetite chromite and barite

Seismic surveys are used extensively to locate salt domes and have been used to a limited extent to prospect for buried channels favorable *for gold* placers or for deposits of uranium Experimental shallow reflection seismic surveys have been made *in connection with prospect drilling to determine* thickness of glaciated material *in the* Lake Supe rior iron region Costs of shallow seismic surveys are comparable with those of electrical resistivity *surveys*

Oxidation of sulfide *ore bodies* or radioactivity associated with extremely rich and large uranium deposits produces heat which may raise the tem perature of the surrounding rocks Temperature measurements from boreholes would therefore seem to offer a usef l pro

rocks a e usually larger than those as ociated with ore bodies

Radioactivity methods Air borne and ground *radioactivity* methods which measure the y radia tion of radioactive elements are used *predomi nantly to* prospect for uranium and thorium They may also be used to prospect for marine *phosphate deposits* which contain small amounts of uranium and for thorium and uranium bearing *placers and beach sands* some of which are mined mainly for titanium minerals Radioactivity logging of bore holes has aided in prospecting for uranium depos its in sandstone by *detecting weakly radioactive* halos that sometimes occur around the ore and in correlating stratigraphic horizons



F g 2 Car-mounted rad oact v ty logg ng equ pment (C M Bunker U S Geolog cal Survey)

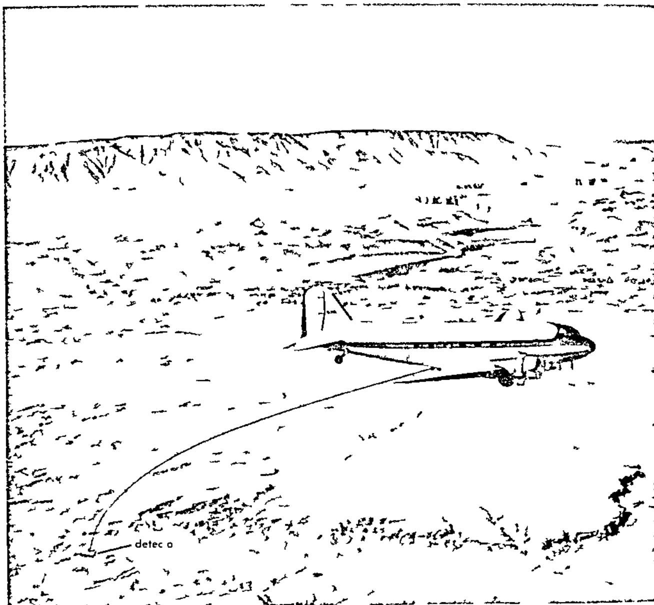


Fig 1 Airborne magnetometer survey over rough terrain (U S Geological Survey)

continuous magnetic formation that may serve as a guide to ore In general a small shallow target requires flying as low as a few hundred feet above the surface whereas larger targets permit flying at higher levels Airborne surveys for magnetic ore deposits at shallow depths are usually made at a more or less constant height above the surface Surveys to outline larger geologic bodies such as buried igneous intrusions or magnetic formations are generally made at a constant barometric altitude

Ground magnetic surveys are best suited to detailed prospecting in small areas often chosen on the basis of aeromagnetic surveys Surveying procedures are determined by the amplitude and extent of the anomalies sought Dip needle and magnetic declination surveys are adequate for delineating high amplitude anomalies such as those associated with near surface magnetic iron deposits Surveys with more sensitive and costly equipment such as the vertical magnetic field balance or the more recently developed nuclear spin magnetometer are generally required to outline anomalies associated with weakly magnetic deposits deeply buried deposits of high magnetism igneous rocks and geologic structures

The applicability of magnetic methods may usually be determined by a knowledge of the habits or arrangement of the deposits sought and by determinations of the magnetization of the pertinent rocks and minerals Hand specimens of strongly magnetic material will deflect a surveyor's compass or dip needle The magnetization of more weakly magnetic material can be estimated by its effect on a sensitive magnetometer or by crushing specimens and separating the magnetic minerals with a hand magnet *See* MAGNETOMETER

Electrical methods Electromagnetic and electrical resistivity surveys are used to locate deposits of metallic sulfides which except for sphalerite are good electrical conductors They may also be used in prospecting for nonmetallic deposits which are generally poor conductors provided there is structural or stratigraphic control and a measur

ble contrast in conductivity The applications are [illegible]verse as each type of deposit presents its own [illegible]roblem

A variety of instruments has been used in elec[illegible]romagnetic prospecting but all measure anoma[illegible]es in the electromagnetic field set up by induced ground currents rather than in the potential dis[illegible]tribution as in resistivity methods. The frequency of the energizing current should be 250–1500 cycles per second Too low frequencies reduce the strength of the induced field but too high frequencies lack depth penetration

A recent development uses natural random alternating magnetic fields of audio and sub audio frequency to locate subsurface electrical conductors such as sulfide ore bodies Distortions of the fields caused by variations in conductivity are measured at several frequencies by use of search coil detectors The new method may be used for air borne or ground surveys Although a high degree of amplification is required theoretically greater lateral and vertical range is possible than with most electromagnetic methods

Air borne and ground electromagnetic surveys have been useful mainly in areas where unaltered sulfide deposits are fairly close to the surface as in the Canadian Shield Other conductors such as fault gouge graphite and some water bearing formations may likewise be indicated by electromagnetic surveys Air borne electromagnetic surveys

[illegible] Depth to flat lying deposits or beds may often be determined fairly accurately using specialized mathematical methods of interpretation The induced potential method is a modified electrical method that shows promise in prospecting for metallic ores

Electrical and [illegible]

[illegible] trodes in boreholes and on the surface and between two or more boreholes have also been successful in locating conducting ore bodies The principles are identical with those governing surface surveys

A spontaneous polarization or so called self potential as large as a volt or so is set up when sulfide ore bodies are oxidized by downward percolating ground water The pattern of currents thus induced may be inexpensively measured by traverses with a potentiometer connected to two nonpolarizable electrodes The accuracy of measurements is reduced by variations in soil type and soil solutions roots of trees and topography which may set up potentials larger than those associated with some ore bodies Local spurious potentials can be eliminated largely by repeating the observations with the electrodes in different positions

Self potential surveys are useless where the deposits are deeper than 100 ft or where they are below the zone of oxidation In the latter case good results have been obtained when the water table is lowered during long dry periods and the sulfide minerals are subject to renewed oxidation

Gravity, seismic and thermal methods Most mineral deposits are too small or have insufficient density contrast to permit use of gravity methods A further disadvantage is the requirement to correct gravity data for the effect of topographic irregularities in mountainous areas Gravity methods have been highly successful in locating buried salt domes with their associated sulfur deposits on the Gulf Coast They have also been used to a minor extent in prospecting for relatively dense deposits such as magnetite chromite and barite

Seismic surveys are used extensively to locate salt domes and have been used to a limited extent to prospect for buried channels favorable for gold placers or for deposits of uranium Experimental shallow reflection seismic surveys have been made in connection with prospect drilling to determine thickness of glaciated material in the Lake Superior iron region Costs of shallow seismic surveys are comparable with those of electrical resistivity surveys.

Oxidation of sulfide ore bodies or radioactivity associated with extremely rich and large uranium deposits produces heat which may raise the temperature of the surrounding rocks Temperature measurements from boreholes would therefore seem to offer a useful prospecting method In practice they have been applied in only a few instances because thermal effects resulting from ground water circulation or conductivity differences in the rocks are usually larger than those associated with ore bodies

Radioactivity methods Air borne and ground radioactivity methods which measure the γ radiation of radioactive elements are used predominantly to prospect for uranium and thorium They may also be used to prospect for marine phosphate deposits which contain small amounts of uranium and for thorium- and uranium bearing placers and beach sands some of which are mined mainly for titanium minerals Radioactivity logging of boreholes has aided in prospecting for uranium deposits in sandstone by detecting weakly radioactive halos that sometimes occur around the ore and in correlating stratigraphic horizons

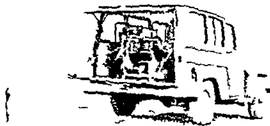


Fig 2 Car mounted radioactivity logging equipment (C M Bunker U S Geological Survey)

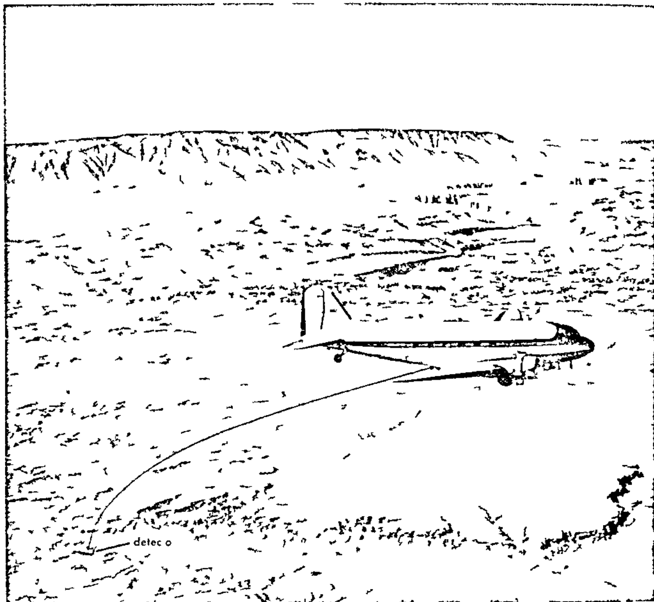


Fig 1 Air borne magnetometer survey over rough terrain (U S Geological Survey)

continuous magnetic formation that may serve as a guide to ore In general a small shallow target requires flying as low as a few hundred feet above the surface whereas larger targets permit flying at higher levels Air borne surveys for magnetic ore deposits at shallow depths are usually made at a more or less constant height above the surface Surveys to outline larger geologic bodies such as buried igneous intrusions or magnetic formations are generally made at a constant barometric altitude

Ground magnetic surveys are best suited to detailed prospecting in small areas often chosen on the basis of aeromagnetic surveys Surveying procedures are determined by the amplitude and extent of the anomalies sought Dip needle and magnetic declination surveys are adequate for delineating high amplitude anomalies such as those associated with near surface magnetic iron deposits Surveys with more sensitive and costly equipment such as the vertical magnetic field balance or the more recently developed nuclear spin magnetometer are generally required to outline anomalies associated with weakly magnetic deposits deeply buried deposits of high magnetism igneous rocks and geologic structures

The applicability of magnetic methods may usually be determined by a knowledge of the habits or arrangement of the deposits sought and by determinations of the magnetization of the pertinent rocks and minerals Hand specimens of strongly magnetic material will deflect a surveyor s compass or dip needle The magnetization of more weakly magnetic material can be estimated by its effect on a sensitive magnetometer or by crushing specimens and separating the magnetic minerals with a hand magnet *See* MAGNETOMETER

Electrical methods Electromagnetic and electrical resistivity surveys are used to locate deposits of metallic sulfides which except for sphalerite are good electrical conductors They may also be used in prospecting for nonmetallic deposits which are generally poor conductors provided there is structural or stratigraphic control and a measur

extent and character of the accumulation discov ered Prospecting methods are the means employed to gain the information called for in these four steps of petroleum prospecting Prospecting meth ods are commonly classified as geological and geo-physical but there is no sharp distinction between the two, all of them involve geological reasoning and interpretation See GEOCHEMICAL PROSPECT ING PROSPECTING

This article outlines two major aspects of petroleum prospecting—geological prospecting and geo physical prospecting

GEOLOGICAL PETROLEUM PROSPECTING

Nearly all prospecting entails certain preliminary library and cartographic background research Some mention of base maps is followed in this section by the topics of surface geology, photo geology, drilling structure and core drilling wild cat wells subsurface geology geological labora tory methods and regional geology

Base maps An essential requisite to petroleum prospecting is accurate base map control Hori zontal control is necessary for location of property boundaries physical features roads wells and other cultural features and for the map location of the points from which geological or geophysical data are obtained Vertical control is necessary for providing topographic information for opera tional purposes as well as for the adjustment of geological geophysical and well data to a com mon datum Topography may also be of important geological significance Aerial photography and electronic and radio positioning systems are largely replacing the theodolite alidade and plane table for mapping and geographic control work both on land and over water and these methods have ad vantages both in speed and in accuracy

Surface geology The examination and study of outcropping rocks as a clue to the structure and stratigraphy of an area is the oldest of petroleum prospecting methods and one which is still ex tremely important The surface geologist maps the topographic expression and distribution of ex posed rock units determines and plots their struc tural attitude measures and describes stratigraphic sections identifies surface structural anomalies such as anticlines or faults which may reflect deeper structures and ...

ing the h ...

ture at ...

surface ...

presence ...

... of petroleum seepages there is no more encouraging indication of a petroliferous province than the presence of actual oil seepages In difficult terrain helicopters com monly attached to surface geological parties aid transportation and communication and facilitate geological observation

Photogeology The mapping of surface geologic features is frequently best carried out through study of aerial photographs In addition to greatly expediting the study of the surface geology of any area the photographic method provides coverage of regions where access on the ground would be prohibitively difficult It usually provides more com plete detail than is possible by surface mapping methods and has the additional advantage of greater over all perspective In regions where out crops are scarce photogeology is employed as a means of determining geologic structure and dis-tribution of formations indirectly through interpre tation of geomorphologic features vegetation frac ture patterns and soil characters Photogeology does not replace surface study which is always desirable but it does constitute an extremely valuable supplement See AERIAL PHOTOGRAPH

Drilling There is no method of petroleum pros pecting so effective as the drilling of a hole to the objective horizon, and if the cost of deep drilling were not so great this method would supplant al ...

...RAL OIL AND GAS WELLS

Structure drilling and core drilling These terms are applied to relatively shallow drilling where the purpose is purely that of securing geological in formation Highly portable drilling rigs with depth capacities of a few hundred to a few thousand feet are used and on the basis of cuttings cores or electrical logs information is obtained on near surface stratigraphy and structure which may guide deeper drilling for petroleum

Wildcat wells Exploratory holes drilled with the aim of discovering new petroleum pools are true wildcat wells Usually these are programmed and equipped for ...

bu ...

per ...

onl ...

equ ...

wa ...

potentially ...

each year be ...

rent method ... the hole is usually kept filled with heavy mud to prevent caving and to hold back excessive fluid pressures consequently po-tentially productive petroleum horizons may fre-quently be penetrated by the bit with very little indication of their fluid content To avoid over looking such horizons rock samples cut by the bit and brought up in the circulating mud are ... and cons ...

instrum ...

rig to a ...

petroleu ... and radioactive devices are run down the hole from time to time to record the properties of the rocks penetrated with respect to the probability of their carrying petroleum Like-wise cores and side wall samples are taken from intervals suspected of being productive

Air borne and ground prospecting is effective only in areas where the radioactive material is exposed at the surface as the γ radiation is effectively masked by a foot or so of barren overburden or a few feet of snow Air borne surveys must be flown within a few hundred feet of the surface because the air similarly absorbs γ rays

Air borne radioactivity surveys could be used to aid in geologic mapping of areas covered by residual soil which usually produces radiation patterns characteristic of the underlying parent rocks In this connection they would be of indirect value in prospecting for nonradioactive minerals There is also evidence that uranium as well as other metals tend to deposit during the last stages of the crystallization of large granitic intrusions This suggests that radioactivity highs over the roofs of large intrusions would be an indirect guide to non radioactive mineral deposits *See* GEOPHYSICAL EXPLORATION

GEOCHEMICAL AND BOTANICAL METHODS

Geochemical prospecting involves the analysis of elements in the soils rocks surface and under ground waters organisms and vegetation for the purpose of defining areas where the distribution of the elements indicates the presence of ore Botanical prospecting involves analysis of elements found in deep rooted plants or plants commonly associated with certain elements or mineral deposits and comparative study of morphological differences in plants growing in mineralized and non mineralized areas

A variety of methods are used singly or in combinations Analysis of residual soils for traces of metals is the most widely used geochemical method as the dispersion pattern corresponds closely with the primary dispersion of metals in the underlying rocks Analyses of metals in vegetation and of the distribution of indicator plants are used in areas where there is little residual soil but the relationship between the metal contents of plants and soil is seldom simple Analysis of water in streams is used for preliminary examinations of large areas Springs usually give information on areas of limited size

The choice of elements for analysis depends on mineral associations relative mobility of the elements and the availability of suitable analytical methods Zinc and copper are useful indicators not only for zinc and copper deposits but for other types of metallic deposits The rare elements in pegmatites and the tin tungsten niobium association in high temperature veins are also useful Associated noneconomic minerals may serve as a guide where the ore metals do not form a recognizable dispersion pattern Immobile metals usually form clear cut dispersion patterns in residual soils whereas mobile metals are more widely dispersed by ground water and vegetation Thus it may be effective to test for the more mobile elements in reconnaissance surveys and the less mobile elements in detailed surveys

Certain bacteria and other organisms concentrate elements such as sulfur selenium boron cobalt and manganese Prospecting for these elements may eventually be aided by additional knowledge of the processes involved and environments required for their concentration *See* GEOCHEMICAL PROSPECTING [H R J]

Bibliography M B Dobrin *Introduction to Geophysical Prospecting*, 1952, A A Fitch D F Christie W E Johnstone and G Whittle Aerial photography in petroleum and mineral prospecting *Empire Mining Met Congr Proc* 4th Congr., part 1 219–247, 1950, S H Dolbear and O Bowles *Industrial Minerals and Rocks* 2d ed 1949, W H Newhouse (ed) *Ore Deposits as Related to Structural Features* 1942

Prospecting, petroleum

The search for commercially valuable accumulations of petroleum This search at the one extreme may be carried out in a completely haphazard manner with entire dependence on luck for success whereas at the other extreme it is a highly organized procedure involving the use of complex precision instruments skilled and experienced personnel and advanced scientific reasoning In either case the final and critical step is always the drilling of an exploratory hole Moreover in neither case can the successful outcome of the exploratory hole be assured in advance because no infallible means of detecting the presence of a commercial petroleum accumulation ahead of the drill has yet been devised Much petroleum has been found both by luck and by the application of scientific methods but statistics demonstrate that at the present time the success ratio of holes located with the benefit of scientific or technical advice is nearly twice as great as that of those located without such advice

The classic requisites for petroleum accumulations are (1) source or mother rocks from which petroleum can have originated (2) carrier and reservoir rocks possessing sufficient permeability to provide avenues of migration as well as sufficient porosity to provide storage space (3) traps adequate to cause commercial concentration of petroleum at local points in the reservoir beds and (4) proper time and spatial relations in the development of source reservoir and trap A favorable hydrodynamic condition might also be mentioned as a requisite to initial accumulation as well as to later preservation of a petroleum deposit *See* PETROLEUM RESERVOIR ENGINEERING

Scientific petroleum prospecting consists of (1) the determination of generally favorable regions with respect to source reservoir trap timing and hydrodynamic conditions (2) the finding of local geological features (anticlines fault traps and pinch outs) within these regions believed to be suited to the trapping of petroleum (3) the location and programming of exploratory holes to test the presence or absence of commercially significant petroleum accumulations on these local features and (4) after initial discovery determination of the

extent and character of the accumulation discov ered Prospecting methods are the means employed to gain the information called for in these four steps of petroleum prospecting. Prospecting meth ods are commonly classified as geological and geo physical but there is no sharp distinction between the two, all of them involve geological reasoning and interpretation *See* GEOCHEMICAL PROSPECT ING PROSPECTING

This article outlines two major aspects of petro leum prospecting—geological prospecting and geo physical prospecting

GEOLOGICAL PETROLEUM PROSPECTING

Nearly all prospecting entails certain preliminary library and cartographic background research Some mention of base maps is followed in this section by the topics of surface geology photo-geology drilling structure and core drilling, wild cat wells, subsurface geology, geological labora tory methods, and regional geology

Base maps An essential requisite to petroleum prospecting is accurate base map control Hori zontal control is necessary for location of property boundaries physical features roads wells, and other cultural features and for the map location of the points from which geological or geophysical data are obtained Vertical control is necessary for providing topographic information for opera tional purposes as well as for the adjustment of geological geophysical and well data to a com mon datum Topography may also be of important geological significance Aerial photography and electronic and radio positioning systems are largely replacing the theodolite alidade and plane table for mapping and geographic control work both on land and over water and these methods have ad vantages both in speed and in accuracy

Surface geology The examination and study of outcropping rocks as a clue to the structure and stratigraphy of an area is the oldest of petroleum prospecting methods and one which is still ex tremely important The surface geologist maps the topographic expression and distribution of [illegible]

[illegible] and prepares cross sections show ing the hypothetical distribution of rocks and struc ture at depth Study of the rocks exposed at the surface may yield important information on the presence and position of source and reservoir rocks as well [illegible]

[illegible]

The [illegible] petroliferous province than the presence of actual oil seepages In difficult terrain helicopters com monly attached to surface geological parties aid transportation and communication and facilitate geological observation

Photogeology. The mapping of surface geologic features is frequently best carried out through study of aerial photographs In addition to greatly expediting the study of the surface geology of any area the photographic method provides coverage of regions where access on the ground would be prohibitively difficult It usually provides more com plete detail than is possible by surface-mapping methods and has the additional advantage of greater over all perspective In regions where out crops are scarce photogeology is employed as a means of determining geologic structure and dis tribution of formations indirectly through interpre tation of geomorphologic features vegetation frac ture patterns and soil characters Photogeology does not replace surface study, which is always desirable but it does constitute an extremely valuable supplement *See* AERIAL PHOTOGRAPH

Drilling There is no method of petroleum pros pecting so effective as the drilling of a hole to the objective horizon and if the cost of deep drilling were not so great this method would supplant al most all others Even so the great bulk of all money spent on petroleum prospecting goes to drilling of exploratory holes *See* BORING AND DRILLING, MIN ERAL [illegible]

[illegible] purpose is purely that of securing geological in formation Highly portable drilling rigs with depth capacities of a few hundred to a few thousand feet are used and on the basis of cuttings cores or electrical logs information is obtained on near surface stratigraphy and structure which may guide deeper drilling for petroleum

Wildcat wells Exploratory holes drilled with the aim of discovering new petroleum pools are true wildcat wells Usually these are programmed and equipped for completion as producers if successful but the so-called [illegible]

[illegible] the hole is usually kept filled with heavy mud to prevent caving and to hold back excessive fluid pressures, consequently po tentially productive petroleum horizons may fre quently be penetrated by the bit with very little indication of their fluid content To avoid over looking such horizons rock samples cut by the bit and brought up in the circulating mud are [illegible] and [illegible] instru [illegible] rig to [illegible] automatically for traces of petroleum, and electrical and radioactive devices are run down the hole from time to time to record the properties of the rocks penetrated with respect to the probability of their carrying petroleum Like wise cores and side wall samples are taken from intervals suspected of being productive

Subsurface geology Regardless of whether production is obtained an exploratory hole is usually a valuable contribution to prospecting knowledge The study of the geological and geophysical data made available through drilling is called subsurface geology The subsurface geologist stationed at the well constantly watches the cores cuttings and drilling fluid for direct traces of petroleum and studies the characteristics of electric logs radioactive logs and geothermal logs for indirect indications of petroleum The lithology paleontology and mineralogy of the cores and cuttings also yield clues to the stratigraphic position at which the well is drilling and the remaining depth to objective producing horizons Determination of the attitude of bedding from cores and from dip-meter surveys gives important evidence as to whether the well is off structure with respect to the fold fault or other trap structure on which it is being drilled and also gives a factor for correcting the drilled thickness of a formation to its true thickness *See* PALEONTOLOGY, PETROLEUM GEOLOGY STRATIGRAPHY

As more wells are drilled in a region the steadily increasing background of subsurface geological information becomes progressively more effective as a means of locating new structural or stratigraphic traps for testing Correlation the identification and tracing of stratigraphic units from one well to another allows conclusions to be reached on the relative structural positions of wells on the probable location of new fold and fault structures on the presence of unconformities and on lateral changes in thickness and lithology These correlation data provide the subsurface geologist with the base for cross sections and various kinds of subsurface maps structure contour isopach paleogeologic lithofacies palinspastic and others *See* FACIES (GEOLOGY) PALEOGEOGRAPHY PALEOGEOLOGY

Geological laboratory methods Many of the determinations which can usefully be made on rock samples either from outcrops or from wells require such specialized knowledge and equipment that the surface or subsurface geologist sends them to specialists in a geological laboratory Paleontologic and micropaleontologic studies are valuable in determination of the age or stratigraphic position of samples in correlation and in determination of past environments of deposition which may bear on source reservoir and stratigraphic trap conclusions Study of the Foraminifera and Ostracoda have been particularly useful in petroleum prospecting and spore and pollen stud es have recently been growing in importance *See* MICROPALEONTOLOGY PALYNOLOGY

Laboratory determination of heavy detrital minerals furnishes useful information for correlation and provenance Among other laboratory methods which may be useful in identification and correlation are analysis for insoluble residues size and shape analysis differential thermal analysis and calcimetry Refractive index determinations made on solvent extracts from rock samples provide useful information on the presence and gravity of even minute traces of oil Computers and data processing machines are being employed in some geological laboratories to aid in the sorting and analysis of large batches of data from surface geology and wells

Regional geology In petroleum prospecting the various contributions of surface geology subsurface geology geophysics and other methods should all be put together and coordinated to give as complete a regional geologic picture as possible Given adequate information on the character and atti [illegible]

dict the location of all of its petroleum accumulations This information is of course never fully forthcoming but the acquisition of as much of it as can be obtained and the imaginative but intelligent extrapolation of the remainder from experience are essential to long range success in petroleum prospecting [H D HE]

GEOPHYSICAL PROSPECTING FOR PETROLEUM

Geophysical techniques have contributed decisively to the world supply of oil since about 1930 The annual cost has amounted to hundreds of millions of dollars during most of these years The seismograph technique has accounted for well over one half of this activity For most of this time and under most circumstances the seismic method has been the most expensive of those available For this reason the cheaper gravity and magnetic methods are often used for reconnaissance purposes and the more limited anomalous areas th is revealed are then subjected to seismic investigation Means and procedures have been developed which permit seismic operations in coastal waters at unit costs comparable to those of gravity and magnetic surveys under the same conditions The rate of expenditure is very high for such operations but the output rate is also high so that acceptable unit costs are achieved The seismic gravity magnetic and the various well logging methods account for all but a tiny fraction of the geophysical work done in the search for oil

The geophysical measurements made in oil prospecting are to a large extent related to the configuration and properties of the rocks which enclose oil pools At best the results of geophysical surveys indicate the presence position and nature of a structure which may or may not contain oil The discovery of the oil itself is made by drilling a hole into the structure If oil is found the well serves the purposes of both discovery and exploitation

Magnetic surveys In the search for oil magnetic surveys are now made almost excl isively from aircraft Oil is found in deep sedimentary basins and the magnetic anomalies found in such areas arise from the igneous floor beneath the sediments The depth below surface to the igneous basement rocks usually is thousands of feet thus a change in flight level by several thousand feet is not criti

cal It is the flight elevation above the igneous surface that counts Aeromagnetic surveys have been made over millions of square miles

The sedimentary structures which are oil bearing often lie above uplifts or topographic features (upper surface irregularities) of the igneous basement surface Local magnetic anomalies are as ... and the slope are a key

of the igneous rocks These are useful in determining the distance below the survey level of the

figuration of the sedimentary basin can be determined and the principal structural features in the basement indicated in advance of any drilling To do this is the principal role of magnetics in oil prospecting

In areas where the sedimentary structure arises mainly from thrusting (force applied sidewise on rockbeds) or is substantially modified thereby magnetic surveys may be of little help See MAGNETOMETER

Gravity surveys These have been most successful in discovering and detailing salt domes a great percentage of which have associated oil accumulation A newly discovered salt dome therefore represents an oil prospect of high potential A salt dome usually contains one or more cubic miles of salt, which is ordinarily of lower density than most of the surrounding sediments A gravity minimum is therefore characteristic of a salt dome structure Salt dome prospects are often drilled on the basis of the gravity anomaly alone

The gravity manifestation of other structural types is generally more complex If a structure involves the position of dense beds nearer to the surface a gravity high will be found Such anomalies are customarily investigated further by seismic techniques before drilling is undertaken

A great percentage of the potential oil producing areas of the United States has been covered by gravity surveys See TERRESTRIAL GRAVITATION

Seismic surveys Contour maps are generally produced to show the elevation of some geological horizon with reference to some datum plane preferably at the general depth level of formations which are known to produce oil elsewhere If there are good seismic reflections above and below this level these are also mapped Together they may furnish evidence of convergence and similar phenomena of practical interest Seismic surveys furnish the most conclusive evidence available from geophysical techniques However areas where there are limestones at the surface or where the sedimentary se...

sto...

oth...

S...

Well logging One or more geophysical techniques of well logging are now applied to practically every well drilled by the oil industry As in pregeophysical days geologists now prepare a graphical log of the formations through which a drill hole extends based on visual examination of drill cuttings brought to the surface by the drilling mud and on core samples Such logs show lithology and fossil distribution with depth Structure maps result from correlations between well horizons which can be identified as being the same or substantially equivalent in all of them Geophysical well logging is merely an extension of this procedure to other physical properties which require physical measurements for their determination

Various geophysical well logging methods have been developed See WELL LOGGING (MINERAL) Acoustic velocity logging may also prove to be a valuable development By this method a continuous record of the variation of the average velocity of sound waves over a short distance is obtained

See GEOPHYSICAL EXPLORATION MINERAL FUEL AREAS [E A F]

Bibliography M S Bishop *Subsurface Mapping* 1960 M B Dobrin *Introduction to Geophysical Prospecting* 2nd ed 1960 J D Haun and L W LeRoy (eds) *Subsurface Geology in Petroleum Exploration* 1958 F H Lahee *Field Geology* 5th ed 1952 K K Landes *Petroleum Geology* 2nd ed 1960 A I Levorsen *Geology of Petroleum* 1954

Prostate disorders

Common prostate disorders include infections calculi and new growths both benign and malignant

Bacterial infections reach the glandular prostate from below by way of the urethra or from above arising from the kidneys ureters or bladder Occasionally infection may be blood borne originating in an infection elsewhere Acute prostatitis is marked by variable fever chills and bladder irritability The prostate is enlarged and painful and abscesses may develop

Chronic prostatitis is often virtually asymptomatic with only a slight discharge vague pain

comes stony hard and nodular In most prostatic infections the adjacent urethra and seminal vesicles are also sites of the infection Prostatitis occurs most commonly in 20 to 35 year old men frequently as the result of gonorrhea or a bladder infection

Calculi or stones are usually precipitated calcium salts that form multiple hard dark concretions They vary considerably in size and range from that of a grain of sand to a marble Calculi often accompany chronic prostatitis or benign prostatic hypertrophy (BPH)

New growths or neoplasia include benign prostatic hypertrophy which occurs in more than one third of men past 60 but may develop earlier Urinary obstruction or a change in urinary habits may be the earliest symptom The gland is diffusely

symmetrically enlarged nontender and compressible The cause is unknown but may be endocrine in nature

Carcinoma of the prostate occurs in about 15% of men over 50 In almost all cases it begins at the periphery of the posterior part of the gland grows slowly and has a variable course Metastases are rare until the gland capsule is invaded then spread may be to the pelvis vertebrae leg bones and regional lymph nodes Ordinarily the first symptom is either local discomfort or urinary obstruction but may sometimes be pain arising in a metastatic site prior to any direct prostatic symptoms

The earliest carcinoma detectable by rectal examination is a small smooth stony hard nodule lying just beneath the capsule Further development results in irregularity and inelasticity but fairly distinct borders remain

A 50% cure rate is obtained following early diagnosis This implies the desirability of a rectal examination at yearly intervals for men past 50 Eighty per cent of the cases that have been undiagnosed until extension beyond the capsule has occurred are inoperable but even here the judicious use of castration estrogen therapy and x ray treatment may prolong life considerably and also give great relief from pain and urinary obstruction *See* ESTROGEN PROSTATE GLAND RADIOLOGY

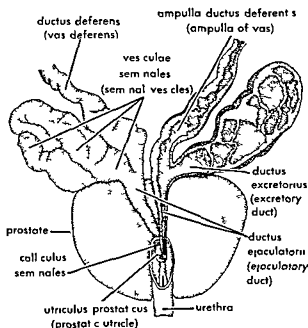
[E G ST]

Prostate gland

In man a triangular body the size and shape of a chestnut that lies immediately in front of the bladder with its apex directed down and forward It is strictly a male organ and has no female counterpart The prostatic portion of the urethra extends through it passing from bladder to penis The organ is composed of 15–20 branched tubular glands which form lobules The gland ducts open into the urethra Between the gland clusters or alveoli there is a dense fibrous supporting tissue the stroma which also forms a tough capsule around the gland and continuous with the bladder wall Penetrating the prostate are the ejaculatory ducts from the seminal vesicles located above and behind the organ which also empty into the urethra The prostatic glands secrete a viscid alkaline fluid which aids in sperm motility and in neutralizing the acidity of the vagina thus enhancing fertilization

A midline groove palpable on rectal examination indicates the two lateral lobes of glandular tissue a median lobe is posterior Small at birth the prostate enlarges to adult size during puberty After middle age the prostate is sometimes subject to new tissue growth usually benign that results in interference with urine flow through the compressed urethra *See* PROSTATE DISORDERS

[E G ST]



Prostate gland and seminal ves cles (After Eycleshymer and Jones in W A N Dorland Am Illus Med cal D ct onary 19th ed Saunders 1942)

Prosthesis

The addition to the human body of some artificial part This article discusses the replacement of living tissue (bone) in the body with a prosthesis Any consideration of the replacement of living tissue with a prosthesis emphasizes the sharp limitations in application For example there is no known substance which can be controlled voluntarily and has the necessary characteristics so that it can be placed within the body to take the place of a muscle Replacement of tendons and ligaments is similarly very difficult with a prosthetic material Here results of at least reasonable success are obtained by transplanting other tendinous or ligamentous tissue to substitute for tendons or ligaments which have been lost In considering endoprostheses that is within the body at the moment one is almost entirely limited to those which replace all or a portion of a bone including its joint surfaces The desirable material for this prosthesis would obviously be the same material as that which is being replaced bone

This in many cases is impossible and would not truly be classified as a prosthesis but rather as a graft Requirements for a prosthetic material to be placed inside the body and to replace bone are as follows (1) it should be nonirritating to the tissues at least to the point where no unpleasant and painful symptoms occur to the patient It is probable that some irritation on a cellular level will be the result of almost any foreign material implanted (2) The material should be literally as strong as possible This does not apply so much to its resistance to a single force but to its fatigue-level life under a corrosive situation which is considerably worse than sea water At the moment there is no known material that has satisfied this qualification This is probably due to the fact that bone is self repairing and therefore literally has no fatigue limit Since bone has this quality of self repair it

is possible to see a comparatively small bony structure surviving repeated stresses over many years. A prosthesis of any known material subjected to the same stresses will undergo fairly rapid fatigue failure (3) Since the prosthesis is replacing only a part of the skeletal system, it must be capable of being coupled to existing bony structures In this field knowledge is quite incomplete In some cases the coupling exists satisfactorily for a number of years if the contact is only that of compression and has a large area of contact However there are many clinical examples of the necrosis and absorption of bone if the area of contact has been too small and therefore the unit loading of the bone too high (4) The contact of the prosthetic material with a joint surface of normal joint cartilage or fibrocartilage which has developed after a shaping of the joint surface by surgical means For a limited amount of function, a satisfactory contact *of the prosthesis and the joint surface can be maintained.*

Application of the prosthesis. *From a review of* the prosthetic characteristics stated above it is evident that *only a* limited amount of function can ever be obtained from the present prosthetic materials applied with the present state of knowledge Under these circumstances one can never expect a complete restoration of normal *function* and prosthetic devices are then reserved for cases in which function is *quite* limited and the prosthetic device is expected only to give a *relatively* improved function, but never to lead to *completely* functional use

Materials for replacing bone The material to be used as a prosthesis replacing bone must have *nonirritating* qualities, the same modulus of elasticity as *that of bone* and the ability to resist corrosion In addition the following problems must be considered coupling of the prostheses *to* living material contact with joint surfaces and *the sites* where prostheses are to be used

1 *Nonirritating qualities of the material* The *nature of* body fluids makes it imperative that the prosthesis *be made of* inert material because of the considerable salt concentration present the oxidation reduction systems which are constantly in action and the *possibility of concentration* cells of either oxygen *or* some of the salt materials which could exist in contact with various portions of the prosthesis It is probable that some tissue reaction could be *generated* by most substances used in the body if they were *present* in an *exceedingly* fine divided state However as solid bits of material as they are used no significant tissue reaction seems to occur from the use *of* type 316 stainless steel the cobalt alloys used under the trademark of Vitallium, and *materials* known as Stellites Glass and some of the plastics are also suitable but lack strength. When using plastics precautions must be taken to be sure that they are completely polymerized since the *monomer* is frequently *toxic* Many plastics however are totally unsuitable for use as prostheses within the body

2. *Strength characteristics* In considering the simple strength characteristics of materials *to* replace bone a desirable feature would be to have material with the same modulus of elasticity as that of bone This would make the problem of the contact between the two materials much simpler The two materials most commonly used type 316 stainless steel and Vitallium, are both about 10 times stiffer than bone At the present time only metals have adequate strength characteristics to replace the structural element of bone and the available metals fall far short of the ideal replacement material.

3 *Fatigue life* The problem of the fatigue life of internal prostheses is a *very complex one* There have been many clinical examples of a brittle failure of metal internal fixation devices which *must* be classed as over-stress since the failure *resulted* after a comparatively small number of cycles As mentioned above the metal is subject to the corrosive action of the salt solutions in the body Oxidation reduction systems are constantly active and *considering the different* blood supplies to which parts of the *prosthesis are exposed* there *must* exist oxygen *concentration* cells and salt concentration cells with their attendant electrolytic action. The problem is *further* complicated by the fact that the prosthesis is *being repeatedly* stressed and is *therefore* subject to the problems related *to* stress *corrosion* An additional complication is found in any prosthesis which is made of two parts since the *infinitesimally* small motions between these *parts may lead* to fretting corrosion with a loss of the protective *oxide* layer This protective coat of [illegible]

[illegible] reoxidize to form a new *protecting layer It* is quite possible however that in the *small places* where contact occurs between *the two* moving parts *of* metal a deficient oxygen concentration may exist with a failure to reestablish the oxide layer *leading* to destructive corrosive action At the *moment* this is perhaps the most serious obstacle *in the way* of producing a true permanent endoprosthesis There is *no* known material with sufficient *fatigue* life under the conditions found in the living body *to* last out even a portion *of the normal* life span if subjected to the stresses of everyday living Our only solution at present is to limit the activities of the patient or to expect at intervals to have to surgically *replace* the prosthesis

4. *Coupling of the prosthesis to living tissue* None of the present-day prosthetic materials can exist in the degree of continuity with living bone which will allow some tension strength Thus they can only be fitted together like bricks without mortar and must depend upon the shape of the contact surfaces for stability There is *no* accurate information available at the present time as to *how much* bone can withstand in repeated *loading* without undergoing necrosis and absorption It has been ob-

symmetrically *enlarged nontender and compressi* ble The cause is unknown but may be endocrine in nature

Carcinoma of the prostate occurs in about 15% of men over 50 In almost all cases it begins at the periphery of the posterior part of the gland grows slowly and has a variable course Metastases are rare until the gland capsule is invaded then spread may be to the pelvis vertebrae leg bones and re gional lymph nodes Ordinarily the first symptom is either local discomfort or urinary obstruction but may sometimes be pain arising in a metastatic site prior to any direct prostatic symptoms

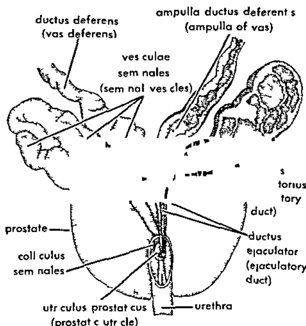
The earliest carcinoma detectable by rectal ex amination is a small smooth stony hard nodule ly ing just beneath the capsule Further development results in irregularity and inelasticity but fairly distinct borders remain

A 50% cure rate is obtained following early di agnosis This implies the desirability of a rectal examination at yearly intervals for men past 50 Eighty per cent of the cases that have been undiag nosed until extension beyond the capsule has oc curred are inoperable but even here the judicious use of castration estrogen therapy and x ray treat ment may prolong life considerably and also give great relief from pain and urinary obstruction *See* ESTROGEN PROSTATE GLAND RADIOLOGY

[E C ST]

Prostate gland

In man a triangular body the size and shape of a chestnut that lies immediately in front of the blad der with its apex directed down and forward It is strictly a male organ and has no female counter part The prostatic *portion of the urethra extends* through it passing from bladder to penis The or



Prostate gland and sem nal vesicles (*After Eycleshymer and Jones in W A N Dorland Am Illus Med cal D ct onary 19th ed Saunders 1942*)

gan is composed of 15–20 branched tubular gland which form lobules The gland ducts open into the urethra Between the gland clusters or alveoli there is a dense fibrous supporting tissue the stroma which also forms a tough capsule around the gland and continuous with the bladder wall Penetrating the prostate are the ejaculatory duct from the seminal vesicles located above and be hind the organ which also empty into the urethra The prostatic glands secrete a viscid alkaline fluid which aids in sperm motility and in neutralizing the acidity of the vagina thus enhancing fertiliza tion

A midline groove palpable on rectal examina tion indicates the two lateral lobes of glandular tissue a median lobe is posterior Small at birth the prostate enlarges to adult size during puberty After middle age the prostate is sometimes sub ject to new tissue growth usually benign that re sults in interference with urine flow through the compressed urethra *See* PROSTATE DISORDERS

[E C ST]

Prosthesis

The addition to the human body of some artificial part This article discusses the replacement of liv ing tissue (bone) in the body with a prosthesis Any consideration of the replacement of living tis sue with a prosthesis emphasizes the sharp limita tions in application For example there is no known substance which can be controlled volun tarily and has the necessary characteristics so that it can be placed within the body to take the place of a muscle Replacement of tendons and ligaments is similarly very difficult with a prosthetic material *Here results of at least reasonable success* are ob tained by transplanting other tendinous or liga mentous tissue to substitute for tendons or liga ments which have been lost In considering endo prostheses that is within the body at the moment one is almost entirely limited to those which re place all or a portion of a bone including its joint surfaces The desirable material for this prosthesis would obviously be the same material as that which *is being replaced bone*

This in many cases is impossible and would not truly be classified as a prosthesis but rather as a graft Requirements for a prosthetic material to be placed inside the body and to replace bone are as follows (1) it should be nonirritating to the tis sues at least to the point where no unpleasant and painful symptoms occur to the patient It is prob able that some irritation on a cellular level will be the result of *almost any foreign material* implanted (2) The material should be literally as strong as possible This does not apply so much to its re sistance to a single force but to its fatigue-level life under a corrosive situation which is consider ably worse than sea water At the moment there is no known material that has satisfied this qualifica tion This is probably due to the fact that bone is self repairing and therefore literally has no fatigue limit Since bone has this quality of self repair it

is possible to see a comparatively small bony structure surviving repeated stresses over many years A prosthesis of any known material subjected to the same stresses will undergo fairly rapid fatigue failure (3) Since the prosthesis is replacing only a part of the skeletal system it must be capable of being coupled to existing bony structures In this field knowledge is quite incomplete In some cases the coupling exists satisfactorily for a number of years if the contact is only that of compression and has a large area of contact However there are many clinical examples of the necrosis and absorption of bone if the area of contact has been too small and therefore the unit loading of the bone too high (4) The contact of the prosthetic material with a joint surface of normal joint cartilage or fibrocartilage which has developed after a shaping of the joint surface by surgical means For a limited amount of function a satisfactory contact of the prosthesis and the joint surface can be maintained

Application of the prosthesis From a review of the prosthetic characteristics stated above it is evident that only a limited amount of function can ever be obtained from the present prosthetic materials applied with the present state of knowledge Unde

tion is quite limited and the prosthetic device is expected only to give a relatively improved function but never to lead to completely functional use

Materials for replacing bone The material to be used as a prosthesis replacing bone must have nonirritating qualities the same modulus of elasticity as that of bone and the ability to resist corrosion In addition the following problems must be considered coupling of the prostheses to living material contact with joint surfaces and the sites where prostheses are to be used

1 *Nonirritating qualities of the material* The nature of body fluids makes it imperative that the prosthesis be made of inert material because of the considerable salt concentration present the oxidation reduction systems which are constantly in action and the possibility of concentration cells of either oxygen or some of the salt materials which could exist in contact with various portions of the prosthesis It is probable that some tissue reaction could be generated by most substances used in the body if they were present in an exceedingly fine divided state However as solid bits of material as they are used no significant tissue reaction seems to occur from the use of type 316 stainless steel the cobalt alloys used under the trademark of Vitallium and materials known as Stellites Glass and some of the plastics are also suitable but lack strength When
taken to b
ized since
plastics h ally unsuitable for use as prostheses within the body

2 *Strength characteristics* In considering the simple strength characteristics of materials to replace bone a desirable feature would be to have material with the same modulus of elasticity as that of bone This would make the problem of the contact between the two materials much simpler The two materials most commonly used type 316 stainless steel and Vitallium are both about 10 times stiffer than bone At the present time only metals have adequate strength characteristics to replace the structural element of bone and the available metals fall far short of the ideal replacement material

3 *Fatigue life* The problem of the fatigue life of internal prostheses is a very complex one There have been many clinical examples of a brittle failure of metal internal fixation devices which must be classed as over stress since the failure resulted after a comparatively small number of cycles As mentioned above the metal is subject to the corrosive action of the salt solutions in the body Oxidation reduction systems are constantly active and considering the different blood supplies to which parts of the prosthesis are exposed there must exist oxygen concentration cells and salt concentration cells with their attendant electrolytic action The problem is further complicated by the fact that the prosthesis is being repeatedly stressed and is therefore stress
cor and in
any s since
the infinitesimally small motions between these parts may lead to fretting corrosion with a loss of the protective oxide layer This protective coat of the metal as far as it is known consists of chrome oxide and in a normal oxygen concentration areas in which the chrome oxide layer have been removed presumably reoxidize to form a new protecting layer It is quite possible however that in the small places where contact occurs between the two moving parts of metal a deficient oxygen concentration may exist with a failure to reestablish the oxide layer leading to destructive corrosive action At the moment this is perhaps the most serious obstacle in the way of producing a true, permanent endoprosthesis There is no known material with sufficient fatigue life under the conditions found in the living body to last out even a portion of the normal life span if subjected to the stresses of everyday living Our only solution at present is to limit the activities of the patient or to expect at intervals to have to surgically replace the prosthesis

4 *Coupling of the prosthesis to living tissue* None of the present-day prosthetic materials can exist in the degree of continuity with
ca
tar
su
tio
bo repeated loading without undergoing necrosis and absorption It has been ob

served clinically, however, that prostheses with relatively small contact area frequently lead to resorption of the bone with a subsequent displacement of the prosthesis On the other hand, the classical example of the prosthesis with a large contact area is the so called Smith Petersen hip cup in which necrosis of the bone is quite rare

The unknown quality in these calculations, however, is the nature of the bone, since the bony structure not only varies greatly in thickness and therefore in strength in the case of osteoporosis, but there is increasing evidence that the structure of the bony substance itself varies to an extreme degree with age and may undergo a progressive loss of blood supply with relatively more and more of the bone becoming avascular and undergoing considerable embrittlement One of the extremely limiting factors in attempting to replace any bone is the space limitation There is only a certain volume of space available beneath the skin of the extremity Most of this is taken up by other important structures, and thus, the prosthesis cannot be significantly larger than the bone which it replaces This is even more critical at the area of contact between the prosthesis and the remainder of the bone Here the size of the prosthesis must be adjusted to the size of the bone although a stem like portion of the prosthesis can be placed inside the marrow cavity of the bone apparently without difficulty It is not possible for the prosthesis to cover the outer surface of the bone to any great extent as well since this will completely shut off its blood supply and eventually lead to its disintegration

5 *Contact with joint surfaces* All substances so far used to replace joint surfaces have been hard and unyielding compared to the considerable flexibility of normal joint cartilage In spite of this considerable success has been experienced in placing a highly polished metal or plastic in contact with either normal joint cartilage or with fibrocartilage which will grow from a denuded bony surface and mold itself to the highly polished prosthesis These new "joints" will stand a considerable amount of function and in a moderate percentage of the cases seem to be quite durable Not all attempts at establishing this new joint are successful however and the factors leading to success or failure are still incompletely understood Once again we find ourselves forced to the conclusion that we do not offer fully normal function but only function to a limited degree

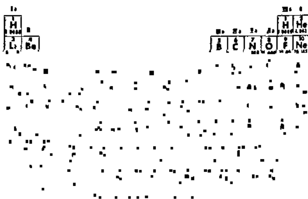
6 *Applications of prostheses* At the present time, replacement of portions of bone has been limited usually to the end of the bone the commonest site is that of the upper end of the femur in the hip joint where attempts have achieved the greatest success

Similar efforts made in almost every joint in the body with varying degrees of success, are usually somewhat more satisfactory in nonweight bearing joints than in weight bearing joints There have been several attempts to replace the central portion of a bone with a metal prosthesis Although some success has been obtained in experimental animals, attempts to use this method in humans has met with almost uniform failure It is possible that the complex stresses imposed upon the tubular shaft of a long bone are too great and lead to failure of the replacement material through fatigue of the contact surfaces between the bone and the prosthesis

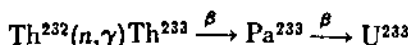
7 *Future possibilities* The eventual thoroughly successful replacement prosthesis must await improvements of material and design encompassing all of the problems discussed above The problem is so complex that a truly satisfactory internal prosthesis cannot be promised within the foreseeable future [C O BE]

Protactinium

Element number 91, protactinium, Pa, is the third member of the actinide group of elements It was discovered by F Soddy and J A Cranston in 1913 and independently, by O Hahn and L Meitner in 1918 *See* ACTINIDE ELEMENTS



All the known isotopes of the element are radioactive The longest lived Pa^{231} is a member of the naturally occurring actinium decay series It decays by the emission of α particles with a half life of about 34 000 years It is present in pitchblende to the extent of perhaps 280 mg/ton in almost the same concentration as radium UX_2 and UZ are short lived, naturally occurring isomeric isotopes of protactinium of mass number 234 The synthetic isotope of mass number 233 is important because it is an intermediary in the production of the fissile isotope of uranium U^{233} from natural thorium Th^{232}, by neutron capture



$$Th^{232}(n,\gamma)Th^{233} \xrightarrow{\beta} Pa^{233} \xrightarrow{\beta} U^{233}$$

Protactinium 233 is β and γ active and decays with a half life of 27 days Several other short lived isotopes with mass numbers between 226 and 237 have been made *See* RADIOACTIVITY

Visible amounts of a compound of the natural Pa^{231} were first isolated by A V Grosse (1927) More recent separation processes applicable to uranium refinery residues generally make use of the extraction of the element from hydrochloric acid solutions by ketones and alcohols such as diisopropyl ketone The protactinium can be reex

solvent These operations can be combined ... cyclic separation and concentration process

Protactinium is similar to the neighboring element thorium in that it does not show a very noticeable resemblance to the heavier actinide elements In its pentavalent state it closely resembles niobium and zirconium The element displays valences of 4 and 5 The pentahalides $PaCl_5$ $PaBr_5$ PaI_5 resemble the niobium and tantalum halides but the pentoxide Pa_2O_5 is much less acidic than the corresponding oxides of niobium and tantalum Compounds of the pentavalent state are very readily hydrolyzed only a few complex anions such as PaF_7 being stable to hydrolysis No stable cationic states have been identified Potassium and ...

... sulfate acid by the addition of hydrogen peroxide Insoluble phosphates and iodates can be precipitated from strongly acidic solutions Many of the pentavalent compounds and most solutions containing this state are colorless However tannic acid precipitates protactinium as a yellow complex from solutions of the soluble complex oxalate Cupferron and pyrogallol also give strong and colored complexes with pentavalent protactinium

Strong reducing agents such as the chromium(II) ion reduce protactinium to the tetravalent state and from such solutions it may be precipitated as the insoluble tetrafluoride The colored tetrahalides and dioxide prepared by dry methods are isomorphous with the corresponding thorium compounds and the other derivatives of this valence state closely resemble the corresponding thorium compounds See NUCLEAR REACTION RARE EARTH ELEMENTS {A G M}

Bibliography G T Seaborg and J J Katz The Actinide Elements Natl Nuclear Energy Series Div IV vol 14A 1954

Protandry

That condition in which an animal is first a male and then becomes a female It occurs in many groups in addition to oysters and cyclostomes The reverse condition is protogyny See PROTOGYNY {T I S}

Proteales

An order of the plant subclass Dicotyledoneae having one family (Proteaceae) with 55 genera and 1200 species It is a dominant group in the drier regions of the Southern Hemisphere About 475 species are South African and 700 are Australian A few occur in South America The silky oak (Grevillea robusta) is grown as an ornamental tree in southern United States See DICOTYLEDONEAE EMBRYOPHYTA PLANT KINGDOM {P D S}

Protective coloration

When the color pattern of an animal increases the probability of its survival it is said to be protectively colored There are three types of protective coloration cryptic or concealing coloration which renders the animal inconspicuous aposematic or warning coloration which advertises the presence of an otherwise well protected animal, and mimicry whereby a species imitates an aposematic species Batesian mimics are edible species which gain protection by mimicking genuine aposematics whereas Mullerian mimicry involves the sharing of a pattern by two aposematic species a type of double protection

Cryptic coloration This phenomenon is widespread The most common form is countershading that is dark pigmentation on the dorsal surface light on the ventral surface An example will explain its utility When a small fish swims near the surface a gull sees the dark dorsal surface blending into the dark waters whereas a larger fish from below sees the light belly blending into the bright surface water Pigmentation is often developed in response to light and countershading may be a simple response to direction of illumination but this does not eliminate its adaptive value

Background matching however may be much more precise and correspondingly more difficult to interpret except as adaptation The peppered moth is represented by two varieties in England A light colored form originally predominant inhabits woods in which the tree trunks are covered by light colored lichens A dark mutant originally rare has become abundant since the industrial revolution It is the common form in the sooty lichen free woods of industrial areas. Both forms on both backgrounds are shown in Fig 1 Studies on ...

... form is taken 10% more often than the dark when adjustment is made for the relative abundance of the two forms Thus protection is genuine and the selective force severe

Background matching may involve morphology and behavior as well as color Thus the walking stick insects of the family Phasmidae closely resemble twigs of the plants upon which they live even to the angle at which they rest upon a branch Numerous insects when at rest resemble the leaves of the plants upon which they feed Kallima, the dead leaf butterfly exemplifies this Background matching may even be actively acquired as in the masking crab, Loxorhynchus which covers its carapace with debris and sessile organisms from its environment If moved to a new location the crab seeks surroundings such as those it left This failing it will remove its covers and replace them with new ones from the new environment

Cryptic coloration may be as useful to predators as to potential prey for it may aid them to approach unseen to make a kill

served clinically, however, that prostheses with relatively small contact area frequently lead to resorption of the bone with a subsequent displacement of the prosthesis On the other hand, the classical example of the prosthesis with a large contact area is the so called Smith Petersen hip cup in which necrosis of the bone is quite rare

The unknown quality in these calculations, however, is the nature of the bone, since the bony structure not only varies greatly in thickness and therefore in strength in the case of osteoporosis, but there is increasing evidence that the structure of the bony substance itself varies to an extreme degree with age, and may undergo a progressive loss of blood supply with relatively more and more of the bone becoming avascular and undergoing considerable embrittlement One of the extremely limiting factors in attempting to replace any bone is the space limitation There is only a certain volume of space available beneath the skin of the extremity Most of this is taken up by other important structures and thus the prosthesis cannot be significantly larger than the bone which it replaces This is even more critical at the area of contact between the prosthesis and the remainder of the bone Here, the size of the prosthesis must be adjusted to the size of the bone, although a stem like portion of the prosthesis can be placed inside the marrow cavity of the bone apparently without difficulty It is not possible for the prosthesis to cover the outer surface of the bone to any great extent as well since this will completely shut off its blood supply and eventually lead to its disintegration

5 *Contact with joint surfaces* All substances so far used to replace joint surfaces have been hard and unyielding compared to the considerable flexibility of normal joint cartilage In spite of this considerable success has been experienced in placing a highly polished metal or plastic in contact with either normal joint cartilage or with fibrocartilage which will grow from a denuded bony surface and mold itself to the highly polished prosthesis These new "joints" will stand a considerable amount of function and in a moderate percentage of the cases seem to be quite durable Not all attempts at establishing this new joint are successful however and the factors leading to success or failure are still incompletely understood Once again we find ourselves forced to the conclusion that we do not offer fully normal function but only function to a limited degree

6 *Applications of prostheses* At the present time, replacement of portions of bone has been limited usually to the end of the bone the commonest site is that of the upper end of the femur in the hip joint where attempts have achieved the greatest success

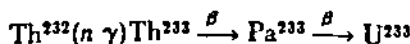
Similar efforts made in almost every joint in the body with varying degrees of success, are usually somewhat more satisfactory in nonweight bearing joints than in weight bearing joints There have been several attempts to replace the central portion of a bone with a metal prosthesis Although some success has been obtained in experimental animals, attempts to use this method in humans has met with almost uniform failure It is possible that the complex stresses imposed upon the tubular shaft of a long bone are too great and lead to failure of the replacement material through fatigue of the contact surfaces between the bone and the prosthesis

7 *Future possibilities* The eventual thoroughly successful replacement prosthesis must await improvements of material and design encompassing all of the problems discussed above The problem is so complex that a truly satisfactory internal prosthesis cannot be promised within the foreseeable future [C O BE]

Protactinium

Element number 91, protactinium, Pa, is the third member of the actinide group of elements It was discovered by F Soddy and J A Cranston in 1913 and independently, by O Hahn and L Meitner in 1918 *See* ACTINIDE ELEMENTS

All the known isotopes of the element are radioactive The longest lived Pa^{231} is a member of the naturally occurring actinium decay series It decays by the emission of α particles with a half life of about 34 000 years It is present in pitchblende to the extent of perhaps 280 mg/ton in almost the same concentration as radium UX_2 and UZ are short lived naturally occurring isomeric isotopes of protactinium of mass number 234 The synthetic isotope of mass number 233 is important because it is an intermediary in the production of the fissile isotope of uranium U^{233} from natural thorium, Th^{232}, by neutron capture



$$Th^{232}(n\ \gamma)Th^{233} \xrightarrow{\beta} Pa^{233} \xrightarrow{\beta} U^{233}$$

Protactinium 233 is β and γ active and decays with a half life of 27 days Several other short lived isotopes with mass numbers between 226 and 237 have been made *See* RADIOACTIVITY

Visible amounts of a compound of the natural Pa^{231} were first isolated by A V Grosse (1927) More recent separation processes applicable to uranium refinery residues, generally make use of the extraction of the element from hydrochloric acid solutions by ketones and alcohols such as diisopropyl ketone The protactinium can be reex



Fig 3 Loxorhynchus the mask ng crab with abundant r ders (From E F R cketts and J Calv n Between Pac fc T des 3d ed Stanford Un versity Press 1952)

Validity of protective coloration Post Darwinian enthus asts interpreted color patterns with uncriti cal exuberance Every dull pattern was considered to be cryptic every gaudy pattern aposematic and every resemblance between species mimicry Evidence that the patterns actually played these roles in nature was not commonly sought Hence during the early 1900s a critical reaction devel oped It was s ggested that colors might not appear to natural predators as they do to man Also di tastefulness of models had not been proven and was open to doubt Finally W L McAtee published a report on the stomach contents of no fewer than 80 000 birds in wh ch he

po

sto

te u oration has eliminated many but numer ous cases require reexamination Many cases have



Fig 4 Meph t s a skunk showing aposematic colora t on (After T I Storer and R L Us nger General Zool ogy 3d ed McGraw-H ll 1957)

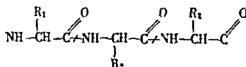
been established experimentally as that of the Af rican birds cited above and the extensive experi ments of Brower on butterflies H B Cott has as sembled a large number of such experimentally verified cases for all types of protective coloration Reexamination of McAtee s data has shown that his statistics are misleading because the numb er of stomachs containing a particular insect was re corded but the number of insects found was not recorded Hence if only a single specimen of a species was taken by a particular bird the record was the same as though 100 had been taken

In conclusion it may be said that protective col oration appears to be a valid evolutionary phenom enon but its importance is not quite so great as was believed in the early post Darwinian era *See* CHROMATOPHORE EVOLUTION ORGANIC [E O D]

Bibliography J v Z Brower Experimental studies of mimicry in some North American butter flies *Evolution* 12 32–47 123–136 and 273–285 1958 H B Cott *Adaptive Coloration in Animals* 1940 E O Dodson *Textbook of Evolution* 1952 W L McAtee Effectiveness in nature of the so-called protective adaptations in the animal king dom chiefly as illustrated by the food habits of Nearactic birds *Smithsonian Inst Publs Misc Coll* 85(7) 1 201 1932

Protein

A high molecular weight compound made up pri marily of a variety of α amino acids joined by peptide linkages involving the α-amino groups and the α-carboxyl groups (*see* AMINO ACIDS)



In some cases the macromolecule consists ex clusively of amino acids (simple protein), in others a nonprotein moiety the prosthetic group is complexed with the protein and forms an integral part of the molecule (conjugated protein) There is no sharply defined line separating the large poly peptides from the small proteins although the term protein is generally reserved for compounds of molecular weight above 3000–5000

Proteins are of central importance in all biologi cal systems playing a wide variety of structural and functional roles They form the primary or ganic basis of such varied structures as hair ten dons muscle, skin and cartilage All of the en zymes the indispensable catalysts of biochemical transformations, are protein in nature Many hor mones are proteins (for example insulin and pro-lactin) or polypeptides (for example adrenocorti cotrophic hormone melanocyte stimulating hor mone oxytocin and vasopressin) The substances responsible for oxygen transport are conjugated proteins containing a metalloporphyrin as the pros thetic group such as hemoglobin and erythrocru orin. The chromosomes warehouses of genetic



Fig 1 Pepper moths of both light and dark varieties photographed on a lichen-covered tree trunk (top) and on a sooty lichen free trunk (bottom) Note blending of each moth into its proper background and the sharp contrast of each to the wrong background (*From Sir Gavin de Beer and Endeavour*)

Aposematic coloration This type of coloration is displayed by animals that are distasteful to predators such as some butterflies those animals equipped with poison glands such as many salamanders and those animals such as bees which have stings or other formidable defenses These may be in sharp contrast to patterns of related animals which are more subject to predation Thus many insects are cryptically colored but bees with formidable stings may be quite gaudy Rodents are generally cryptically colored but the porcupine with its effective armor of quills presents a striking black and white pattern Similarly the skunk which repels almost all predators presents an arresting pattern It cannot be assumed however that a pattern is aposematic simply because it is gaudy The pattern of the tiger is an extraordinarily obtrusive one in captivity Yet, in its native habitat, the alternating black and yellow stripes blend into the deep shadows and the intense highlights of the tropical jungle This then, is a cryptic pattern which facilitates close approach to the prey

Mimicry. This phenomenon is widely distributed but has been best studied in the Lepidoptera The experiments of J Brower on American butterflies are most illustrative Caged scrub jays were used as predators Nonmimetic butterflies were eaten whenever offered Models that is, noxious species were consistently refused after a preliminary experience Mimics were generally taken by birds which had not been conditioned to the models but were usually refused by birds which had been so conditioned The famous case of the monarch and viceroy butterflies has generally been treated as Batesian mimicry, but the viceroy was taken by unconditioned birds less readily than were nonmimetic butterflies, hence this example may be intermediate between the two kinds of mimicry *Battus philenor*, a black butterfly is consistently rejected by the jays *Papilio troilus P polyxenes* and *P glaucus* are similar species which are rejected by conditioned but not by unconditioned jays The last however, is not rejected as consistently as the others, and so may be distinguished from the model Among birds the African drongo *Dicrurus* is solid black and is inedible A flycatcher, *Bradyornis ater* and the cuckoo shrike, *Campephaga nigra* are edible but mimic the drongo The tit *Parus niger*, is solid black ventrally but has prominent white markings dorsally When freshly killed specimens of all three were offered ventral surface up to a cat all were refused When the same birds were offered dorsal surface up the tit was quickly taken although the others were refused

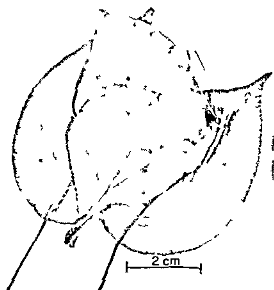


Fig 2 *Kallima* the dead leaf butterfly photographed with two leaves from its habitat (*General Biological Supply House Inc*)

the polypeptide chain The pitch of the helix is such that 3 7 amino acids are contained in one complete turn There is no doubt of the importance of the helical configuration in fibrous proteins There is also evidence for a similar orientation of the polypeptide chains in globular proteins but here the data are not as clear cut In addition to hydrogen bonds there are electrostatic forces such as those between —COO and —NH_3^+ groups in side chains and Van der Waals forces which help determine the configuration of the polypeptide chain The term secondary structure is used to refer to all those structural features of the polypeptide chain determined by noncovalent bonding

respect to the others as in collagen and this inter helix structure can be referred to as tertiary structure

Denaturation Many of the characteristic properties of a protein depend upon the integrity of the secondary structure which is considerably less stable than the primary covalently bonded structure The biological activity of enzymes and of protein hormones for example can be destroyed by mild procedures which do not break any of the covalent bonds Heating exposure to weak acid or alkali, solution in the presence of high concentrations of certain salts are all procedures that can profoundly alter secondary structure without splitting peptide bonds, and any such disorienting process is called denaturation The most apparent effect of denaturation is a decrease in solubility Heat coagulation of proteins such as those of egg white is a familiar example of denaturation Other accompaniments of denaturation include an increase in the reactivity of the amino acid side chains and changes in optical rotation The latter effect depends upon the fact that the coiling of the polypeptide chain confers optical activity on the protein superimposed on that due to the asymmetric carbon atoms in its component amino acids As more and more of the helical configuration is lost there is a progressive increase ...

readi

Iso

... properties of the protein to be isolated and the nature of the source material used The decided instability of most proteins imposes limitations on the conditions that can be employed during isolation if denaturation is to be avoided In the case of a biologically active protein it is possible to use its enzymatic or hormonal activity as an index of alterations occurring during isolation When no such index is available it is often impossible to be certain that the product finally isolated retains the same molecular properties that as its form within the living cell In order to minimize the chances of denaturation, it is best to observe certain general rules derived from experience with known proteins (1) Temperatures should be kept as low as possible This is particularly important when organic solvents such as ethanol and acetone are used (2) Extremes of pH are to be avoided (3) High concentrations of protein are preferable since most proteins are relatively more stable when concentrated (4) Excessive agitation and foaming are to be avoided since proteins denature more rapidly at surfaces and interfaces

Mechanical disruption of the tissue or cell source is usually necessary for efficient extraction of the proteins The tools used for this disruption can vary from a meat grinder to a Waring Blendor to an ultrasonic oscillator Once in solution any combination of the methods described below may be applied

Salting out High concentrations of neutral salt tend to precipitate proteins, the concentration necessary for such precipitation varying widely from one protein to another The most commonly used salt for the purpose is ammonium sulfate which has the advantages of high solubility at the low temperatures useful in protein fractionation and a low temperature coefficient of solubility It is conventional to record the concentration of ammonium sulfate as the percentage of complete saturation at the given temperature rather than as the absolute molar concentration By adding salt stepwise to a protein solution and centrifuging off the precipitate at each step a series of arbitrary fractions is obtained in one or two of which the desired protein will generally be concentrated For example the addition of ammonium sulfate to serum to 34% of saturation precipitates most of the γ globulins and very little of the other protein components This 34% "cut" can then be centrifuged ...

... effects of salts on protein solubility are better described in terms of ionic strength defined by the following equation

$$\text{ionic strength } (\Gamma/2) = \tfrac{1}{2}\sum C_i Z_i^2$$

C_i denotes the molar concentration of a given ionic species in the solution Z_i denotes the charge on that species and the sum is taken over all species ...

... ration can be approximated by the following simple relationship derived by E J Cohn

$$\log S = \beta - K_s(\Gamma/2)$$

where S = solubility in grams per liter, K_s = a salting out constant dependent on the nature of the protein and the salt used but independent of pH

information in the cell, are highly complex nucleo proteins that is, proteins conjugated with nucleic acid The viruses are also nucleoprotein in nature Even this incomplete catalogue makes clear the ubiquity and fundamental significance of proteins in the life processes *See* HEMOGLOBIN

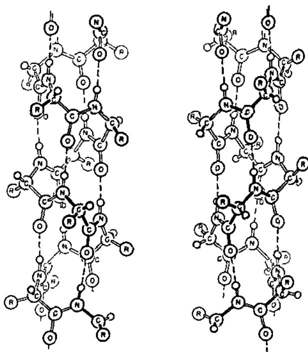
Structure The proteins are essentially polymers of α amino acids joined in peptide linkage Only α amino acids of the L configuration are found in proteins and approximately 20 different amino acid species have been recognized as common con stituents of protein glycine, alanine, valine, leu cine, isoleucine serine, threonine cysteine, cystine, methionine aspartic acid, asparagine glutamic acid glutamine lysine arginine, histidine phenyl alanine, tyrosine tryptophan, proline They occur in widely varying proportions in different proteins and some proteins are completely lacking in one or more of these In addition several amino acids occur only in certain highly specific proteins Hy droxyproline, for example has been found only in collagen and elastin connective tissue proteins of animal tissues *See* articles on individual amino acids

The properties of a protein are determined in part by its amino acid composition For example the net charge on the macromolecule at any given hydrogen ion concentration is largely a function of the relative number of dibasic and dicarboxylic amino acids This net charge strongly influences the solubility of the protein at different pH values since the solubility depends in part on the propor tion of polar groupings on the macromolecule When the hydrogen ion concentration is high (low pH) the net charge is positive, when the hydrogen ion concentration is low (high pH), the net charge is negative The pH at which the net charge of the *protein is zero is defined as its isoelectric point and* solubility is at a minimum at this pH

The sequence in which the different amino acids are linked in any given pure protein is highly specific and characteristic for that protein This specificity of sequence is one of the most remark able aspects of protein chemistry The number of possible permutations of sequence in even so small a protein as insulin (molecular weight 5732, 51 amino acid residues) is astronomic—10^{51} permu tations Yet it is now rather well established that the cell actually makes only one of these possible sequences The elucidation of the mechanism con ferring such a high degree of specificity on the biosynthetic reactions by which proteins are built up from free amino acids is one of the key problems of modern biochemistry

In addition to the backbone of α peptide bonds the amino acids in the polypeptide chain or chains of a protein may be covalently linked in other ways The most important of these is the disulfide bridge joining two cysteine residues at different points along the same polypeptide chain or two cysteine residues in otherwise distinct polypeptide chains This is known as cross linkage Other modes of cross linkage have been discovered such as phos phate diester bonds in pepsin, but these are de cidedly less common The arrangement of the amino acids in the peptide chains together with the other structural features determined by cova lent linkages, is known as the primary structure of the protein molecule

A single chain of several hundred amino acids could assume an almost infinite number of different spatial configurations This multiplicity of con figurational possibilities is due to the free rotation about covalent bonds (other than those in the pep tide bonds, which are rigid) Largely as a result of x ray diffraction studies, it is now known that the long polypeptide chains of proteins particularly those of the fibrous proteins, are in fact held in a rather well defined configuration (*see* PROTEIN FIBROUS) The backbone is coiled in a regular fashion forming an extended helix As a result of this coiling, peptide bonds separated from one an other by several amino acid residues are brought into close spatial approximation The stability of the helical configuration is attributable to the for mation of hydrogen bonds between these peptide bonds The particular helical configuration which best fits the available x ray data is that proposed by Linus Pauling and R B Corey In this structure each —NH group in peptide bond is hydrogen bonded to the $-\overset{O}{\overset{\|}{C}}-$ group 3 residues beyond it in



The α helix of Pauling and Corey The repeating —N—C—C— units form the backbone which spirals up in the left handed or a right handed fashion Hy drogen bonds are indicated by the dotted lines Note that the side chains (R) are all directed out from the helix The pitch of the helix is about 5 4 A, and there are 3 7 residues contained in one complete turn (From J T Edsall and J Wyman, Biophysical Chemistry vol 1, Academic Press, 1958)

tion of polypeptides than in fractionation of proteins See EXTRACTION

Immune precipitation If a protein has already been obtained in pure form a specific antibody against it can be prepared by repeated injection of the pure material into a laboratory animal The serum of this animal containing the specific antibody can then be used as a specific precipitating agent for the removal of the antigenic protein from mixtures The limitations of the method lie in the relatively frequent occurrence of cross reactions the difficulty in dissociating the antibody antigen complex and the practical difficulties of large scale application See PRECIPITIN REACTION

Preparative ultracentrifugation Proteins of high molecular weight can be concentrated at the bottom of a centrifuge tube by applying a high gravitational field for a prolonged time Clearly it is possible to achieve significant purification only for the heaviest or the lightest components in a mixture and the degree of purification will depend on the spectrum of sedimentation velocities represented Proteins with densities less than that of the medium can be made to float to the surface under the influence of a strong gravitational field The latter principle has been applied with signal success in the study of serum lipoproteins which by virtue of their lipid content have densities below those of other serum proteins Sufficient salt is added to the serum to raise its density above that of any of the lipoproteins yet not above that of the other serum proteins Centrifugation to equilibrium (100 000 × g for 20 hr) brings the lipoproteins to the surface where they can be collected by aspiration or by actually slicing through a celluloid tube to separate the upper few milliliters from the infranatant solution This technique has been refined by the use of repeated centrifugation steps at progressively rising salt densities so that subfractions of the lipoproteins in different density classes can be obtained on a preparative scale See ULTRACENTRIFUGE

Preparative electrophoresis In the classical free electrophoresis system of A Tiselius the mixture of proteins being analyzed occupies one segment of the U shaped vessel When the current is applied the fastest migrating component moves out ahead of the other proteins in the mixture and the slowest migrating component lags behind the body of proteins in the mixture Thus the proteins of maximum and minimum mobility can be isolated in relatively pure form but those of intermediate mobility can be only partially purified by this procedure Therefore modified methods more suitable for preparative purposes have been introduced.

In zone electrophoresis the protein mixture is applied as a very narrow band either on filter paper a block of starch or some other supporting medium wetted with an appropriate buffer solution When current is passed through the supporting medium each protein moves out from the line of origin with characteristic velocity and in a given time with a given current flow migrates a characteristic distance from the line of origin Thus at the end of the run the proteins of the mixture are separated in group fashion from one another The supporting medium can then be divided into a number of separate zones, and the proteins in each zone can be separately washed off the paper or starch for analysis and further purification

Convection electrophoresis a method devised to permit larger scale preparation of proteins combines separation in the horizontal direction by electrophoresis with separation in the vertical direction by convective transport The protein mixture is introduced into a tall narrow vessel fitted at the bottom with a large reservoir When an electric field is applied normal to the vessel s narrow dimension the protein of highest mobility tends to concentrate against one wall and then to sink down toward the reservoir by convection Ultimately then there is progressive concentration of this protein in the bottom reservoir The material concentrating in the reservoir can in turn be subjected to a repetition of the procedure and thus it is possible to obtain considerable purification of rather large amounts of protein See ELECTROPHORESIS

Criteria of purity Because of their high molecular weight and their relative instability the proteins do not lend themselves to characterization by the classical methods of organic chemistry Aside from providing information about prosthetic groups elementary analysis is not helpful An important exception occurs in the case of certain conjugated proteins that include a stoichiometric amount of a metal for example the cytochromes

[illegible]

are not obtainable and decomposition points are not sharp Crystallinity once thought to be rather good proof of purity is now well recognized to be inconclusive Frequently a crystalline preparation although very rich in the desired material proves upon careful analysis by more sensitive methods to be a mixture of two or more proteins It has been necessary to develop special methods for protein characterization and many of these are closely related to methods used for isolation of pure proteins

The most important generalization to be noted is that none of the methods commonly used will establish positively the purity of a protein A preparation apparently pure by one criterion may well prove to be inhomogeneous by another The [illegible] that [illegible]

by a[illegible]

and [illegible]

d[illegible]res [illegible] judging the purity of a protein (see COLLOID) If the material migrates as a single component in an electric field and sediments as a single component in a centrifugal field this is strong evidence of purity Examination at different ionic strengths and pH values may reveal inhomogeneity in a protein apparently pure under a given set of conditions What actually is demonstrated is that

and temperature, and β = logarithm of the hypothetical solubility at zero ionic strength a value strongly dependent on pH and temperature

With certain sparingly soluble proteins such as myosin purification is better achieved by extracting away the more readily soluble proteins leaving the desired material in the residue in partially purified form

Isoelectric point precipitation At any given salt concentration protein solubility varies with pH and as already mentioned is at a minimum at the isoelectric point By raising the salt concentration and adjusting the pH to the isoelectric point it is often possible to obtain a precipitate considerably enriched in the desired material In some cases the desired component can be crystallized directly from a heterogeneous mixture of proteins by this simple procedure Ovalbumin for example can be obtained in crystalline form from egg white in essentially two steps First the globulins are precipitated by adding ammonium sulfate to 40% of saturation The supernatant is then adjusted to pH 4 7 the isoelectric point of ovalbumin and the ammonium sulfate concentration is gradually increased until a slight opalescence develops Addition of seed crystals is frequently all that is necessary to initiate crystallization It is important to note that protein crystals obtained in this way are rarely free of contaminant proteins Repeated recrystallization or application of other purification procedures is required to obtain a truly homogeneous protein preparation

Dialysis and dilution Some proteins the true globulins are relatively insoluble at very low ionic strengths and can be precipitated from a mixture simply by dialyzing against water so as to progressively remove salt from the solution Alternatively a large volume of water may be added to reduce ionic strength and thus effect precipitation

Precipitation with organic solvents Organic solvents such as ethanol and acetone sharply reduce the solubility of most proteins At room temperature denaturation proceeds very rapidly in the presence of these reagents but they can be and have been very successfully employed by carrying out all operations at temperatures near or below 0°C The outstanding example of their use is in the systematic separation of the components of human plasma by methods developed at the Harvard Biophysical Laboratory by E J Cohn John T Edsall and their co workers By appropriate selection of pH ionic strength temperature and ethanol concentration a series of precipitated fractions can be obtained which although mixtures represent considerable degrees of purification of each of the major groups of serum proteins This group separation can then be followed by subfractionation steps leading to progressively purer components

Formation of protein salts and other complexes On the acid side of the isoelectric point proteins act as polyvalent cations and can form salts with many anions that reduce their solubility The commonest reagents for nonspecific precipitation of proteins (trichloroacetate phosphotungstate, perchlorate picrate) act in this way The anion can be added in acid form thus simultaneously lowering the pH and effecting salt formation so that the proteins immediately precipitate out These precipitants irreversibly denature most proteins but they are extremely useful for separating protein from nonprotein components when further purification is not required and denaturation is irrelevant Heavy metal cations are also effective protein precipitants Under appropriate conditions advantage may be taken of the reduced solubility of metal salts of proteins to improve the selectivity of protein fractionation The crystallization of insulin as the zinc salt may be cited as an example The use of zinc and barium in plasma protein fractionation reduces the concentration of ethanol to which they must be exposed and improves selectivity Interaction of proteins with other large molecules (proteins nucleic acids polysaccharides) may also be useful in purification procedures

Adsorption Materials such as aluminum oxide aluminum silicate calcium phosphate gel and kaolin will adsorb proteins from solution and exhibit some degree of selectivity in this adsorption under appropriate conditions The desired component can be adsorbed onto the solid phase and subsequently eluted or conversely the adsorbent solid phase can be used to remove contaminating proteins leaving the desired component in solution

Paper chromatography This has not yet been widely used for purification of proteins but it has been signally successful in a few instances such as in the purification of insulin from pancreatic extracts on a small scale It has been a key method on the other hand in the separation of peptides in connection with studies of protein structure *See* CHROMATOGRAPHY

Column chromatography This procedure using ion exchange resins and chemically modified cellulose derivatives can be a powerful tool Low molecular weight proteins especially those with markedly acidic or basic isoelectric points such as ribonuclease histones and lysozyme can be separated from most other proteins in a mixture in a single run on an ion exchange column Powdered cellulose chemically treated to introduce charged groups is particularly suitable for protein fractionation since the conditions of pH and ionic strength used are in a range where most proteins are stable By the use of eluants graded continuously in pH and ionic strength remarkable discrimination can be achieved and the capacity of the columns permits the handling of gram quantities of protein

Countercurrent distribution When two mutually immiscible solvent systems are available countercurrent distribution is an effective system for purification but the range of application has been narrow The power of the method is shown by the successful separation by Lyman Craig of insulin into two components differing only by a single amide grouping As with chromatography this method has been much more widely utilized in fractiona

T Svedberg equation

$$M = \frac{RTs}{D(1 - V\rho)}$$

where R is the gas constant T is the absolute temperature and s is the sedimentation constant The latter expresses the rate of migration of the protein under unit centrifugal acceleration

If centrifugation is carried out at a considerably lower rotational velocity the sedimentation of the protein molecules will eventually be exactly counterbalanced by the back diffusion from the area of higher concentration From the concentration pattern at this equilibrium the molecular weight can be calculated This sedimentation equilibrium method suffers the disadvantage that several days of centrifugation may be necessary to achieve true *equilibrium* A theoretical treatment of sedimentation equilibrium due to Archibald now makes it possible to determine molecular weight even for very small molecules with a very short centrifuge run The fundamental basis for this analysis is the fact that equil brium is rapidly established at the meniscus and at the bottom of the cell From a *measurement* of solute concentration and concentration gradient at these two points in the cell it becomes po sible to calculate molecular weights Separate measurement of the diffusion constant is not necessary

Osmotic pressure Using membranes of collodion or cellophane which are permeable to small molecules but impermeable to *proteins it is possible* to determine molecular weights by measuring equilibrium osmotic pressures across the membranes Even very concentrated solutions of *proteins of* high molecular weight yield relatively low osmotic pressures and the method is therefore *not* easily appl ed to proteins with molecular weight above 100 000 For the study of proteins with low molecular weight it is necessary to prepare special membranes of suitable pore size and this makes application difficult below molecular weights of *about* 5000 Nevertheless it has been possible to determine accurately the molecular weight of the insulin *monomer (6000)* by this method

Light scattering and low-angle x ray scattering The scattering of light incident on a solution *is a* function of the number and size of the solute molecules This method as applied to *proteins* is considerably more flexible than the osmotic pressure method and can be used over a much wider range of molecular weights X rays by virtue of their very short *wavelength* are scattered at much smaller angles relative to the primary beam but the same theoretical bases apply

Other methods Information regarding the shape of protein molecules can be obtained by measurement of a number of other physical properties including double refraction of flow actual size as determined by electron microscopy viscosity rotatory diffusion coefficient and relaxation time Details of the theoretical and practical aspects of these and other physical methods can be found in the excellent review article by Ed all

Classification In 1907 a combined committee representing the American Society of Biological Chemists and the American Physiological Society proposed a formal classification of proteins into three major categories (1) simple proteins containing only amino acids (2) conjugated proteins containing a covalently bonded nonprotein constituent in addition to amino acids (3) derived proteins products resulting from treatment of proteins with such agents as heat alcohol acid and enzymes The last category embraces all denatured proteins and hydrolytic products of protein break

water but soluble in neutral salt solutions

c Glutelins insoluble in all neutral solvents but soluble in dilute acid or alkali

d Prolamines insoluble in water *absolute* ethanol and *other neutral solvents* but *soluble in* 70–80% ethanol

e Albuminoids highly insoluble in all *neutral* solvents

f Histones *soluble in water* insoluble in ammonia *and* coagulated by heating Rich in basic amino acids *but* containing a wide variety of other *amino acids*

g Protamines soluble in water not coagulated by heat Contain almost *exclusively* the basic amino acids.

Over the years it has become clear that the above *classification* is not completely satisfactory and by

o large groups the globular proteins soluble in aqueous or ethanolic *solutions* and the fibrous proteins *structural* proteins insoluble *in ordinary* aqueous media The latter correspond roughly to the *albuminoids* or scleroproteins of earlier classifications *See* ALBUMIN GLOBULIN GLUTELIN PEPTIDE PROLAMINE PROTEIN GLOBULAR

[D.ST]

Bibliography M L Anson K Bailey and J T Edsall (eds) *Advances in Protein Chemistry* 1944–1959 E J Cohn and J T Edsall

veurath and K Bailey (eds) *The Proteins* vols 1 and 2 1953–1954 H D Springall *The Structural Chemistry of Proteins* 1954

all of the proteins in the preparation have the same or very similar charge and molecular size and shape Subtle differences in structure will not be detected Insulin for example behaves as a homogeneous protein when examined by these methods but by the use of countercurrent distribution it is shown to be a mixture of two components which differ only by one amide group It should be noted that impurities amounting to less than 5% of the total are detectable only with great difficulty by electrophoresis or ultracentrifugation

One of the most sensitive criteria of purity depends upon application of Gibbs phase rule Increasing amounts of protein are added to a series of vessels containing a constant amount of solvent The vessels are gently shaken until equilibrium is reached and then the amount of protein in solution is measured If the protein is pure the plot of the amount of protein dissolved against the amount added to the vessel will be a straight line function up to the saturation point and then break sharply to a plateau value If there are two or more proteins in the preparation there will be two distinct breaks in the curve or there will be no sharp inflection point at all

The techniques of immunochemistry can be applied in several ways to the study of protein homogeneity (*see* IMMUNOLOGY) Antigen antibody reactions while highly specific are not completely so Consequently the most sensitive methods utilize a combination of immunochemical properties and physical properties For example antibody prepared against a protein preparation can be incorporated into a gel and the solution being tested for homogeneity can then be allowed to diffuse through the gel If the preparation is pure a single zone of precipitation will appear at the point in the advancing front where antibody antigen ratio is optimal for the precipitin reaction If more than one antigenic protein is present more than one band will appear at points dependent on the diffusion constant of the proteins and on their immunochemical properties Another highly sensitive method immunoelectrophoresis combines zone electrophoresis with immune precipitation

Methods end groups and amino acid sequence A simple polypeptide chain will have a single free α amino group at one end and a single free α carboxyl group at the other The amino acid residues bearing these uncombined groups are designated respectively the N terminal residue and the C terminal residue and a number of methods are available for specifically identifying these end groups The most widely used method for determining N terminal residues introduced by F Sanger in 1945 depends on reaction of the protein with fluorodinitrobenzene The N terminal residues are converted to their α N dinitrophenyl derivatives and after acid hydrolysis of the protein these can be readily identified by suitable extraction and chromatography A widely used enzymatic method for determining the C-terminal residues is the reaction of the protein with carboxypeptidase which specifically splits off residues bearing a free α carboxyl group Alternatively the proteins can be reacted with anhydrous hydrazine which breaks up the peptide chains and converts the internal residues in the protein to hydrazides The C terminal residues are released as free amino acids and can be identified as such in the mixture If a protein gives a stoichiometric yield of a specific end group or end groups this is strong supportive chemical evidence of purity Of course this does not prove that the remainder of the structure is homogeneous but it can rule out contamination with proteins bearing end groups different from those of the major component

Some of the most important advances in protein chemistry in the period from 1945 to 1958 lay in the exploitation of methods for determining amino acid sequence and other features of the primary structure Insulin was the first protein for which the complete structure was elucidated and Sanger received the 1958 Nobel Prize in Chemistry for his classical work leading to that structure No general method is yet available that permits stepwise degradation of a protein residue by residue although beginnings have been made in that direction Instead it is necessary to partially degrade the protein using pure proteolytic enzymes which split at specific peptide bonds to yield a mixture of smaller peptides Each of these can in turn be attacked by a combination of end group methods amino acid analysis and some of the stepwise degradative procedures By combining the results of a variety of such procedures it is then possible to reconstruct the sequence within each peptide Finally by combining the results obtained by degrading the original protein with proteolytic enzymes of different bond specificity the order of the peptide fragments in the protein can be deduced The location of disulfide cross linkages can also be established by modern methods and this has been accomplished for insulin and for ribonuclease Again the approach is to degrade the protein with a proteolytic enzyme which now yields some peptides containing disulfide bridges (cystine residues) These can be located on paper chromatograms and their structure determined By comparison of the amino acid sequences adjacent to the half cystine residues with the over all structure of the protein the points along the chain or chains connected by disulfide bridges can be identified

Methods size and shape of proteins Methods for the determination of the size and shape of proteins are presented in the following sections

Sedimentation and diffusion The molecular weight of a protein is most commonly determined from a measurement of its rate of sedimentation in a centrifugal field In addition to the rate of sedimentation one must know also the partial specific volume (V) the density of the solvent (ρ) and the diffusion constant of the protein (D) The molecular weight can then be calculated from the

more of the character of a simple extended poly peptide chain like the fibrous proteins

Most of the data on which the L. Pauling and R B Corey α helix theory of protein structure is based derive from the study of fibrous proteins Recently direct x ray crystallographic evidence has been added to the available indirect evidence that the polypeptide chains of the globular proteins are also at least in part in α helical configuration and that these helices are in turn folded in complex fashion to yield the compact shape of the native protein *See* X RAY CRYSTALLOGRAPHY [D ST]

Protein metabolism

The transformation and fate of food proteins from their ingestion to the elimination of their excretion products Proteins are of exceptional importance to organisms because they are the chief constitu ents aside from water of all the soft tissue of the body Special proteins have unique roles as struc tural and functional elements of cells and tissues Examples are keratin of skin collagen of tendons actin and myosin of muscle the blood proteins enzymes in all tissues and protein hormones of the hypophysis *See* BLOOD ENZYME HORMONE MUSCLE

Isotopic labeling experiments have established that body proteins are in a dynamic state con stantly being broken down and replaced This is a rapid process in organs active in metabolism such as liver kidney intestinal mucosa and pancreas much slower in skeletal muscle and extremely slow in connective tissue elements and skin

Protein is digested to amino acids in the gastro intestinal tract These are absorbed and distributed among the different tissues where they form a series of amino acid pools that are kept equili brated with each other through the medium of the circulating blood The needs for protein synthesis of the different organs are supplied from these pools Excess amino acids in the tissue pools lose their nitrogen by a combinat on [illegible]

[illegible] to the other major foodstuffs—car bohydrates and fats *See* CARBOHYDRATE LIPID

The recommended daily protein intake is 1 g/kg body weight for adults, for example 70 g for a 165 lb man Th [illegible]
up to 1 ye [illegible]
children ar [illegible]
pregnanc [illegible]
m [illegible]
ne [illegible]
[illegible] body protein and as a source of various other body constituents derived from the amino acids. The amino acids of proteins fall into two nutritional categories essential or indispensa ble and nonessential or dispensable For a number of amino acids the category to which they belong changes between the periods of body growth and adulthood and in different animal species

Table 1 Classification of amino acids with respect to growth effect in white rat[a]

Essential or indispensable		Nonessential or dispensable	
Lysine	Isoleucine	Glycine[c]	Hydroxy proline
Tryptophan	Methionine	Alanine	Citrulline
Histidine	Valine	Serine	Cystine[d]
Phenyl alanine	Threonine	Aspartic acid	Tyrosine[e]
Leucine	Arginine[b]	Glutamic acid	
		Proline	

[a] After W C Rose *Phys Rev* 18 109 1938
[b] Arginine can be synthesized by the rat, but not at a sufficiently rapid rate to meet the demands of normal growth
[c] Glyc ne a equent [illegible]

The nutritional classification of the amino acids for the growth of the rat is shown in Table 1 It has been found to hold generally for a number of car nivores and omnivores Such measurements have not been made on growing children Ruminants synthesize practically all the amino acids through the action of the bacteria of the rumen

The essential amino acids for maintenance of nitrogen equilibrium in healthy young men and the daily requirement is given in Table 2 This list com prises only eight amino acids The remaining amino acids can be formed in the body from other ma terials. Only in this sense can the dietary dispensa ble amino acids be considered nonessential All of the constituent amino acids are essential for pro tein formation and certain of them are the pre cursors of such important body substances as crea tine thyroxin adrenalin histamine, and the pu rines and porphyrins *See* AMINO ACIDS.

It might be expected that in conditions of aug mented protein need such as pregnancy and lacta tion and after trauma or in specific pathological conditions certain of the dispensable amino acids would become indispensable due to overtaxing of the synthetic capacity There is almost no informa tion on this point In the disease phenylketonuria it has been indicated that tyrosine becomes indis pensable because there is a block in the conversion of phenylalanine to tyrosine.

Table 2 Essential amino acids for normal man when the diet furnishes sufficient nitrogen for the synthesis of the nonessentials*

Amino acid	Minimum daily requirement g/day	Recom mended daily intake † g/day	Number of subjects tested
L-Tryptophan	0 25	0 5	37
L Phenylalanine	1 10	2 2	28
L-Lysine	0 80	1 6	33
L-Threonine	0 50	1 0	24
L-Valine	0 80	1 6	29
L-Methionine	1 10	2 2	19
L-Leucine	1 10	2 2	14
L Isoleucine	0 70	1 4	14

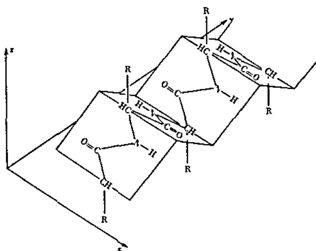
* After W C Rose *Chem Eng News* 30.2385 1952
† These figures represent in each case a safe intake and are not to be regarded as optimum

Protein, fibrous

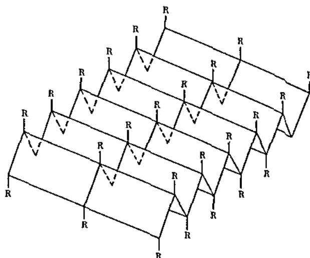
A general name for a group of highly insoluble proteins which form the principal structural basis of many animal tissues Examples of fibrous proteins are keratin of hair, horn and feathers, myosin of muscle, fibrinogen the plasma protein responsible for the structure of the clot, collagen of tendon, cartilage and fish scales None of the ordinary solvents will solubilize these proteins Collagen, when treated with boiling water, is converted to gelatin, ordinary household glues are essentially hot water extracts of animal connective tissue *See* FIBRINOGEN, GELATIN

The fibrous proteins have been extensively studied by means of x ray diffraction and the information gained from these studies forms the foundation for many of the modern concepts of polypeptide structure Two general types of x ray diffraction pattern are observed One is common to keratin myosin and fibrinogen and this group has been designated the kmf group of fibrous proteins The collagens from various sources give closely related patterns which are quite distinct from those of the kmf group *See* X RAY DIFFRACTION

X ray diffraction studies reveal only periodicities of structure These must then be interpreted within the framework of chemical evidence regarding the structure of the protein For example a 3 33 A spacing parallel to the fiber axis of keratin in the extended form (β) is demonstrated by x ray studies In order to explain this spacing, which is shorter than that expected for the residue length in a fully extended polypeptide L Pauling R B Corey and H R Branson have proposed that the

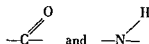


Pleated ribbon The peptide chain backbone is pleated parallel to the *x* axis at each α carbon The identity distance *along the y axis is 6 68 A* corresponding to a distance of 3 33 A per amino acid residue (*From H D Springall The Structural Chemistry of Proteins Academic Press, 1954*)



Pleated sheet Three pleated ribbons are shown joined side to side by lateral hydrogen bonding The inter chain distance is 4 73 A (*From H D Springall, The Structural Chemistry of Proteins Academic Press, 1954*)

bon above is held to an adjacent ribbon by hydrogen bonding between



groups in the peptide bonds a pleated sheet is formed The calculated spacing between adjacent ribbons would be 4 73 A, in good agreement with the x ray findings

An important characteristic of the keratins is their ability to reversibly extend and contract Studies of the α form reveal quite different spacings It is largely from analysis of the patterns of *fibrous proteins in the α-form that Pauling and Corey developed their proposal of the 3 7 residue* helix (*see* PROTEIN) It should be noted that the question of the ubiquity of this structure particularly the question of its importance in globular proteins, is not yet resolved [D ST]

Protein, globular

The term for a broad class of proteins readily soluble in ordinary aqueous solvents as opposed to the fibrous proteins which are generally insoluble in water dilute acid or dilute alkali (*see* PROTEIN FIBROUS) Included in globular proteins are the great majority of proteins in the animal organism (such as hemoglobin ovalbumin casein insulin and virtually all enzyme proteins and serum proteins) For many of these proteins the solubility properties appear to depend upon an organized three dimensional arrangement of folded polypeptide chains When this organized structure is disrupted by denaturation using any of several chemical or physical procedures the solubility decreases markedly (*see* PROTEIN) Furthermore there is x ray crystallographic evidence that denatured globular proteins have some of the structural features of fibrous proteins Denaturation appears to represent an unfolding of the molecule

more of the character of a simple extended poly peptide chain like the fibrous proteins

Most of the data on which the L. Pauling and R B Corey α helix theory of protein structure is based derive from the study of fibrous proteins ... evidence has ... idence that ... proteins are also at least in part in α helical ... guration and that these helices are in turn folded in complex fashion to yield the compact shape of the native protein *See* X RAY CRYSTALLOGRAPHY [D ST]

Protein metabolism

The transformation and fate of food proteins from their ingestion to the elimination of their excretion products Proteins are of exceptional importance to organisms because they are the chief constituents aside from water of all the soft tissue of the body Special proteins have unique roles as structural and functional elements of cells and tissues Examples are keratin of skin collagen of tendons actin and myosin of muscle the blood proteins enzymes in all tissues and protein hormones of the hypophysis *See* BLOOD ENZYME HORMONE MUSCLE.

Isotopic labeling experiments have established that body proteins are in a dynamic state constantly be ng broken ...

... These are absorbed and distributed among the different tissues where they form a series of amino acid pools that are kept equilibrated with each other through the medium of the circulating blood The needs for protein synthesis of the different organs are supplied from these pools Excess amino acids in the tissue pools lose their nitrogen by a combination of transamination and deamination The nitrogen is largely converted to urea and excreted in the urine The residual carbon products are then further metabolized by pathways common to the other major foodstuffs—carbohydrates and fats *See* CARBOHYDRATE LIPID.

The recommended daily protein intake is 1 g/kg body weight for adults, for example 70 g for a 165 lb man This is increased to 3 5 g/kg for infants up to 1 year and from 1 5 to 2 g/kg for growing children and for women during the latter half of pregnancy and during nursing

Role in diet Ingestion of protein is needed primarily to supply amino acids for the formation of new and depleted body protein and as a source of various other body constituents derived from the amino acids The amino acids of proteins fall into two nutritional categories essential or indispensable and nonessential or dispensable For a number of amino acids the category to which they belong changes between the periods of body growth and adulthood and in different animal species

Table 1 Classification of amino acids with respect to growth effect in white rats[a]

Essential or indispensable		Nonessential or dispensable	
Lysine	Isoleucine	Glycine[c]	Hydroxy proline
Tryptophan	Methionine	Alanine	Citrulline
Histidine	Valine	Serine	Cystine[d]
Phenyl alanine	Threonine	Aspartic acid	Tyrosine[d]
Leucine	Arginine[b]	Glutamic acid	
		Proline	

[a] After W C Rose *Phys Rev*, 18 109 1938

[b] Arginine can be synthesized by the rat but not at a sufficiently rapid rate to meet the demands of normal growth

[c] Glycine is essential for the growing chick

[d] When adequate amounts of these amino acids are available in the diet, the requirement for methionine and phenylalanine respectively is diminished

The nutritional classification of the amino acids for the growth of the rat is shown in Table 1 It has been found to hold generally for a number of carnivores and omnivores Such measurements have not been made on growing children Ruminants synthesize practically all the amino acids through the action of the bacteria of the rumen

The essential amino acids for maintenance of nitrogen e... daily requ... prises onl... acids can be formed in the body from other materials Only in this sense can the dietary dispensable amino acids be considered nonessential All of the constituent amino acids are essential for protein formation and certain of them are the precursors of such important body substances as creatine, thyroxin adrenalin, histamine and the purines and porphyrins *See* AMINO ACIDS

It might be expected that in conditions of augmented protein need such as pregnancy and lactation and after trauma or in specific pathological conditions certain of the dispensable amino acids would become indispensable due to overtaxing of the synthetic capacity There is almost no information on this point In the disease phenylketonuria, it has been indicated that tyrosine becomes indispensable because there is a block in the conversion of phenylalanine to tyrosine

Table 2 Essential amino acids for normal man when the diet furnishes sufficient nitrogen for the synthesis of the nonessentials[*]

Amino acid	Minimum daily requirement, g day	Recommended daily intake †, g/day	Number of subjects tested
L-Tryptophan	0 25	0 5	37
L-Phenylalan ne	1 10	2 2	29
L-Lysine	0 80	1 6	33
L-Threonine	0 50	1 0	24
L-Valine	0 80	1 6	29
L-Methionine	1 10	2 2	19
L-Leucine	1 10	2 2	14
L-Isoleucine	0 70	1 4	14

[*] After W C Rose *Chem Eng News* 30 2385 1952

† These figures represent in each case a safe intake and are not to be regarded as optimum

Nutritive value of proteins All proteins are not equally nutritious Animal proteins are generally superior to vegetable proteins Rarely this may result from resistance to digestion which usually is counteracted by cooking or heating A well known example is soybean meal The nutritive value of its protein is improved by heating because this destroys a substance in the meal which inhibits digestion by trypsin Overheating lowers the nutritional value of proteins by making lysine unavailable This is a problem of some concern in connection with the manufacture of prepared breakfast cereals (*see* FOOD ENGINEERING) The major cause of poor nutritional value however is a low content or unavailability of one or more of the indispensable amino acids Vegetable proteins tend to be lacking in lysine and tryptophan

DIGESTION OF PROTEIN

This occurs to a limited extent in the stomach and is completed in the duodenum of the small intestine The main proteolytic enzyme of the stomach is pepsin which is secreted by the chief cells in an inactive form pepsinogen Its transformation to the active pepsin is initiated by the acidity of the gastric juice and accelerated and completed by pepsin The activation process involves liberation of a portion of the pepsinogen molecule as a *peptide Pepsin preferentially hydrolyzes peptide* bonds containing an aromatic amino acid and it requires an acid medium to function *See* DIGESTIVE SYSTEM, PEPSIN

A second proteinase in the stomach rennin present only in infancy is particularly adapted to the digestion of milk protein Digestion is initiated by the well known milk clotting reaction used in cheese manufacture Rennin requires less acid than pepsin to be active In infancy hydrochloric acid secretion by the stomach is not fully developed *See* CHEESE RENNIN

Digestion in intestine The acid chyme is discharged from the stomach containing partially degraded proteins into a slightly alkaline fluid in the small intestine This fluid is composed of pancreatic juice and succus entericus the intestinal secretion The pancreas secretes three known proteinases trypsin chymotrypsin and carboxypeptidase All three are secreted as inactive zymogens Activation starts through the action of a substance present in the intestinal secretion itself a specific enzyme—enterokinase This transforms the inac*tive trypsinogen into the active trypsin* This conversion also is hastened by the autocatalytic activity of trypsin Trypsin in turn activates chymotrypsin and carboxypeptidase In all of these activation processes certain peptide bonds are broken to yield the active enzymes The mucosa of the small intestine contains various peptidases which are not liberated into the intestinal fluid but apparently act by contact at the cell surface or by absorption of the split products produced during intestinal digestion *See* PEPTIDE

Trypsin and chymotrypsin are endopeptidases that is they cleave internal peptide bonds The so-called peptidases are exopeptidases They cleave terminal peptide bonds Trypsin has a predilection for those containing the basic amino acid residues of lysine and arginine These two proteinases perform the major share in hydrolyzing proteins to small peptides Digestion to amino acids is completed by the exopeptidases Carboxypeptidase acts on peptides from the free carboxyl end aminopeptidases from the free amino end Other peptidases act on di or tripeptides or peptides containing such special amino acids as proline

The absorbed amino acids are carried by the portal blood system to the liver From there they are distributed to the rest of the body

The amino acid digestion products of the proteins are absorbed as rapidly as they are liberated The absorption is confined chiefly to the small intestine and is a process that involves the metabolic participation of the cells of the intestinal mucosa Small amounts of the peptides formed during digestion escape further hydrolysis and may also enter the circulation from the intestine This is shown by a rise in the peptide nitrogen in the blood

The permeability of the intestinal mucosa for undigested protein appears greater in infancy This in combination with the low concentration of *digestive enzymes appears responsible for the im*munological sensitization often observed in infants particularly for milk and egg proteins Thus the digestion of protein is necessary not only to yield small diffusable compounds that are readily absorbed from the intestine but also to eliminate the antigenic properties of proteins which could produce harmful allergic reactions

To serve the needs for protein synthesis all the constituent amino acids must be introduced into the body simultaneously Withholding of an indispensable amino acid even for a few hours produces growth retardation or a negative nitrogen balance

Protein in feces formation The unabsorbed food residue in the small intestine is passed into the cecum then the colon and finally is eliminated as feces Water is absorbed from the liquid mass leading to a more solid consistency in the cecum and ascending colon The fecal material is composed of undigested food residues bile pigments leukocytes bacteria and the products of secretion of the intestine and pancreas—enzymes mucus and desquamated epithelial cells The protein present in the feces comes from the above sources as much as one fourth of the dried feces may consist of bacteria These bacteria also act on the amino acids liberated during digestion and produce degradation products useless for the metabolic needs of the body The most conspicuous of these are indole and skatole formed from tryptophan which are chiefly responsible for the odor of feces Roughly 1 g of nitrogen per day is carried by the feces largely present in the bacteria

UTILIZATION OF ABSORBED AMINO ACIDS

For tissue protein synthesis The absorbed amino acids that escape decomposition become part of the amino acid pools of the body From these amino acids new tissue proteins are synthesized to meet body needs The need for new tissue protein is *greatest in childhood* during growth and in adults after protein depletion following fasting or con valescence from a wasting or debilitating disease This is associated with a positive nitrogen balance *leading to* an increase in body nitrogen

In addition turnover of tissue proteins occurs in the adult animal in nitrogen balance with no net gain of body nitrogen This is demonstrated by isotopic tracer experiments and has led to the hypothesis that the body proteins are continually undergoing synthesis and degradation but remain relatively constant in quantity The rate of replace ment varies greatly for different tissues In man it has been estimated that the average half life of the total body protein is 80 days, that of lung brain

protein is supported by observations on the differ ence in protein loss by the tissues of the body in a 7 day fast by the rat The liver lost 40% of its *protein the alimentary tract* pancreas and spleen

in determining the protein nutritional status of the individual A blood sample is easily drawn and estimation of the different plasma proteins is now becoming standard procedure The plasma proteins are quite labile and show marked fluctuations *in* conditions associated with a disturbance of protein *metabolism*

The major organ of plasma protein synthesis is *the liver* It forms all of the plasma albumin and fibrinogen and a considerable proportion of the globulins Advanced liver disease results in hypo albuminemia and a lowered fibrinogen content Prolonged protein deprivation both diminishes the albumin content and causes damage to the hepatic cells *See* ALBUMIN FIBRINOGEN

A portion of the total plasma globulin is syn thesized in other tissues containing reticuloendo thelial cells *The hormones* and enzymes present in blood plasma are derived in the main from non hepatic sources

The plasma *proteins* have numerous important physiological functions The albumin is the major factor in the regulation of the blood volume through its osmotic action which counteracts the fluid expulsion *effect of the hydrostatic* pressure resulting from the contractions of the heart Fi brinogen is only one component of a sequential process essential for coagulation of the blood Other plasma components include the blood plate lets and prothrombin The globulins include frac tions that are carriers of phospholipids and sterols and certain essential metal ions iron and copper Other fractions, chiefly γ globulin contain the anti bodies that are the defenses against numerous diseases

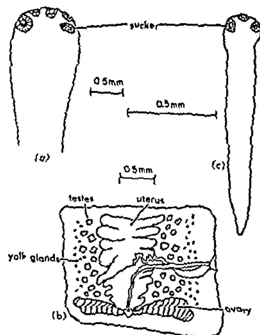
Synthesis and utilization of the plasma proteins is a rapid process Much of the knowledge of this has been learned from studies *on the rate of re newal of the plasma proteins and the albumin* in particular in health and in disease by isotopic labeling methods These studies have shown that there is a complete turnover of the major plasma proteins in a period of a few days The *difference* from normal in the turnover times in a variety of diseases provides an insight *into the nature* of the disease processes [D M G]

Bibliography F R McHenry

Proteocephaloidea

An order of tapeworms of the subclass Cestoda With one exception these worms are *intestinal* parasites of fresh water fishes amphibians, and *reptiles The holdfast* organ bears four suckers and frequently, an apical organ which may be

within the central mesenchyme of the *segment and the* Monticelliidae in which some or all of the organs are in *the cortical* mesenchyme and which parasitize catfishes The life histories of



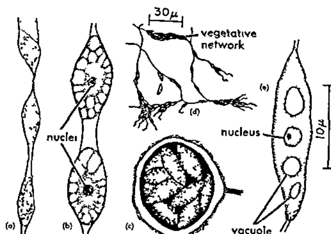
Proteocephalus (a) Scolex (b) Segment (c) Plerocer coid larva

several proteocephalids have been studied After leaving the host's intestine, the embryos must be ingested by an arthropod, typically a copepod of the genus *Cyclops*, in the body cavity of which the young worms grow, developing into procercoids and finally plerocercoids (illustration c) The ver*tebrate host becomes infected by eating Cyclops* containing plerocercoid larvae

Proteocephalid like forms in the ancient fresh water fishes probably were ancestral to the other orders of Cestoda *See* CESTODA, *see also* TETRAPHYLLIDEA [C P R]

Proteomyxida

This is a poorly defined group of the subphylum Sarcodina The Proteomyxida or Proteomyxa are considered to constitute a class of the Rhizopodea Three families have been assigned to this group (1) *Labyrinthulidae (see illustration)* are uninucle



Labyrinthulidae (a) *Labyrinthula zopfi* portion of living network (after Valkanov) (b) *L zopfi*, two organisms stained (after Valkanov) (c) *L zopfi*, encysted stage (after Valkanov) (d) *L macrocystis*, vegetative network (after Cienkowski) (e) *L macrocystis*, single organism stained (after Cienkowski) (From R P Hall, *Protozoology*, Prentice-Hall 1953)

ate marine organisms found on eel grass and certain algae The organisms secrete filaments along which they glide An aggregate of many individuals may form a motile pseudoplasmodium *Labyrinthula macrocystis* has been found on eel grass suffering from fungal disease (2) Pseudosporidae are nonplasmodial types which may invade filamentous algae and Volvocidae (3) Vampyrellidae are plasmodia at maturity Some species (*Vampyrella*) may invade algae by digesting the cell wall, certain others (*Leptomyxa*) may be secondary *invaders in diseased hops See* RHIZOPODEA [R P H]

Proterostomia

That part of the animal kingdom in which cleavage of the egg is of the determinate type The mesoderm originates from certain cells or cell masses and the blastopore becomes the mouth This group presumably derivable from a trochophorelike larva, includes all bilateral phyla except the Echinodermata, Chaetognatha, and Chordata *See* DEUTEROSTOMIA [T I S]

Proterozoic

The name for the interval of geological time between the Archeozoic, or early Precambrian time and the Cambrian Period of the Paleozoic Era A

PRE CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
						CARBONIFEROUS							
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	*Mississippian*	*Pennsylvanian*	PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY

great interval of land erosion or unconformity separates the Proterozoic strata from the underlying Archean rocks Some of the most important characteristics of Proterozoic formations are (1) their partly undeformed or less deformed condition compared with older strata, (2) the abundance of clastic sediments and interbedded lava flows of the Keweenawan type in many places, (3) the occurrence of highly hematitic iron formation in some regions, and (4) the presence of sediments of glacial and interglacial origin in places *See* PRECAMBRIAN, UNCONFORMITY [M E W]

Bibliography A L DuToit *The Geology of South Africa* 3d ed 1954, W G Wilmarth, *Geologic Time Classification of Geological Survey Compared with Other Classifications*, USGS Bull 769 1925

Proteus

A genus of gram negative bacteria of the family Enterobacteriaceae (see ENTEROBACTERIACEAE) *Members of the genus* are characterized by the possession of enzymes decomposing proteins and urea, and by high motility by means of peritrichous flagella This enables *Proteus* to spread or "swarm" over the surface of solid culture media Four species are usually distinguished *Pr vulgaris*, *Pr mirabilis*, *Pr morgani*, and *Pr rettgeri* The bacteria are found in putrefying matter soil feces of man and animals and are frequently a cause of urinary tract infections These microorganisms are also found in abscesses and contaminated wounds, and may be involved in infant diarrhea (*see* ENTERIC BACILLI)

Serological analysis by means of the somatic (O) and the flagellar (H) antigens permits subclassification into numerous serotypes (*see* SALMONELLA) Cross reactivity of some of these types with rickettsiae is utilized in important diagnostic

procedures for typhus (Weil Felix reaction) and other rickettsial diseases *See* BACTERIOLOGY MEDICAL, IMViC TEST, RICKETTSIOSES [A J W]

Protista

A kingdom of organisms proposed by E H Haeckel in 1866 to include all unicellular organisms previously grouped under the phylum Thallophyta of the plant kingdom and under the phylum Protozoa of the animal kingdom Twelve groups of problematical single celled organisms are also included, whose affinities with the plants or animals are not readily apparent Organisms of the kingdom Protista lack definite cellular arrangement Although they may be colonial in habit all cells are identical, and there is no differentiation of cells for specific purposes or functions Representative organisms classed under the Protista include the bacteria molds fungi algae diatoms, infusorians foraminifers and radiolarians *See* MICROPALEONTOLOGY, PROTOZOA THALLOPHYTA [D J J]

Bibliography R C Moore Kingdom of organisms named Protista *J Paleontology* 28(5) 588–598 1954

Protobranchia

A small and primitive subclass in the class Pelecypoda the bivalve mollusks In all families other than the *Solemyidae* the hinge is taxodont that is the numerous hinge teeth are uniform vertical and have corresponding sockets in each valve There is also a central ligament pit The anterior and posterior adductor muscles are nearly equal in size In most groups the extensions of the labial palps are greatly developed for feeding These emerge from the valves and collect food material on the ocean floor The gills are used only for respiration and with included muscle fibers function as a pumping membrane, drawing water from the inhalant chamber to the exhalant chamber

This subclass is economically important only as a source of food for bottom feeding fish All are marine animals and many are found at great depths *See* PELECYPODA [W J C]

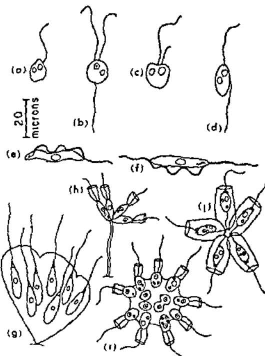
Protogyny

A condition in hermaphroditic or dioecious animals in which the female reproductive structures mature before the male structures It is of rare occurrence Botanically protogyny occurs in some plant species in which the stigma develops withers and dies before the anthers mature [T I S]

Protomastigida

An order in the phylum Protozoa containing a heterogeneous group of colorless flagellates possessing one or two flagella in some stage of their life cycle These small organisms (5–80 μ in length) typically have pliable bodies Some species are holozoic and ingest solid particles while others are saprozoic and obtain their nutrition by absorption Their life cycles are usually simple but some species have two or more recognizably distinct stages The species may be either free living or parasites of vertebrates invertebrates and plants Reproduction is by longitudinal fission although multiple fission occurs in some species Several are important disease producing parasites of man and of domestic and wild animals

Taxonomy. The order Protomastigida includes all of the Protozoa with only one or two flagella that are not markedly ameboid do not contain chloroplasts and are not closely related to chloroplast bearing flagellates Organisms that possess one or two flagella and are decidedly ameboid belong to the family Mastigamoebidae in the order Rhizomastigida Those which are colorless but considered to be close relatives of protozoans possessing



Representative genera of families of order Protomastigida (a) Oikomonas (family Oikomonadidae), one anterior flagellum (b) Amphimonas (family Amphimonadidae) two equally long anterior flagella (c) Monas (family Monadidae) two unequally long anterior flagella (d) *Bodo* (family Bodonidae), two unequally long flagella one of them trailing (e) *Trypanosoma* (family Trypanosomatidae) one flagellum with undulating membrane (f) *Cryptobia* (family Cryptobiidae) two flagella one free and one with undulating membrane (g) *Phalansterium* (family Phalansteriidae) several individual cells each with one flagellum and narrow collar embedded in gelatinous substance (h) *Codosiga* (family Codosigidae) several individual cells each with one flagellum and a distinct collar clustered on the end of a stalk (i) *Protospongia* (family Codosigidae), individual cells embedded in gelatinous substance, the outer cells each with a collar and flagellum (j) Bicosoeca (family Bicosoecidae) clustered individual cells each with a flagellum and small collar contained within vaselike lorica

chloroplasts belong to the family Chlamydomonadidae of the order *Phytomonadida*

There is not good agreement on the division of the order into families However the five or more families can be divided into three general groups The first group contains simple organisms with no distinctive features save one or two flagella of equal or unequal length This includes the families Oikomonadidae Amphimonadidae Monadidae and Bodonidae (illustration *a d*) The second group contains organisms which in addition to one or two flagella have an undulating membrane the families included in this group are Trypanosomatidae and Cryptobiidae (illustration *e* and *f*) The third group includes the organisms possessing a peculiar collar surrounding a single flagellum it includes the families Phalansteriidae Codosigidae and Bicosoecidae (illustration *g j*) These collared flagellates are referred to as choanoflagellates and constitute an interesting group of free living Protozoa The collar which is retractile appears to be a tubular extension of the protoplasm It apparently assists the flagellum in directing food particles into an area of the body where they can be ingested

Choanoflagellates The colonial choanoflagellate *Protospongia* (illustration *i*) produces a gelatinous substance with flagellated collared individuals on the outer surface and unadorned ameboid members embedded in the interior This organism is regarded as a possible link between the unicellular Protozoa and the parazoan sponges Sponges have cells which have flagella surrounded by collars lining internal cavities Since these particular structures are known to occur in only these two groups of organisms it is believed that they may be related

Some of the choanoflagellates are retained in a goblet or vase shaped lorica with a long stalk which may attach either to the substratum or to the lorica of another organism

Trypanosomatidae The most important family of the Protomastigida is the Trypanosomatidae since it includes several species that infect man and his domestic animals with serious diseases such as African sleeping sickness The organisms in this family are polymorphic changing their form in various stages of their development Their life cycles may involve one or two hosts (invertebrate vertebrate or plant) The trypanosome form possesses a single flagellum and an undulating membrane extending the full length of the body The other related forms are simpler lacking one or both of these structures *See* MASTIGOPHORA PORIFERA TRYPANOSOMATIDAE [M M B]

Bibliography R P Hall *Protozoology* 1953
T L Jahn and F F Jahn *How to Know Protozoa* 1949

Proton

An elementary particle which is the positively charged constituent of ordinary matter Together with the neutron the proton is the building stone of all atomic nuclei a single proton constitutes the nucleus of the hydrogen atom The most important properties that characterize the proton are its charge which is identical in magnitude but of opposite sign to that of the electron (the negatively charged constituent of ordinary matter) and has the value of 4.8029×10^{-10} esu $= 1.6021 \times 10^{-19}$ coulomb its mass 1.6724×10^{-24} g $= 1836.1 m_e$ (m_e is the mass of the electron) its spin $\frac{1}{2}\hbar = \frac{1}{2}\, 1.0544 \times 10^{-27}$ erg sec ($\hbar$ is Planck s constant h divided by 2π), its magnetic moment 1.4104×10^{-23} erg/gauss its lifetime which according to all available evidence is infinite and the fact that it obeys the Pauli exclusion principle that is it is a fermion (obeys Fermi Dirac statistics) *See* ELEMENTARY PARTICLE

Mass and charge measurements The determination of the mass and charge of the proton can be made with different degrees of precision in many ways Deflection of proton beams in electric and magnetic fields gives the ratio of the charge to the mass of the proton The fact that the hydrogen atom is neutral guarantees the equality in absolute value of the charge of the electron and of the proton and thus every method for measuring the charge of the electron such as the celebrated oil drop experiment of R A Millikan (1909) also gives the charge of the proton Spectroscopic observations also contribute to this determination mostly giving the ratio between the charge and mass of the electron Moreover all methods which give Avogadro s number indirectly give in addition the mass of the proton (*see* ATOMIC CONSTANTS) Indeed the strict connection and interdependence of many lines of approach to such quantities as the charge and mass of the proton are among the strongest and most striking supports of the modern theories of atomic and nuclear physics

Range of protons in matter Protons of high velocity lose their energy in matter by several mechanisms Occasionally they strike another nucleus and then may be elastically scattered or may produce a nuclear reaction These events drastically alter the proton energy and if the proton is in a beam it is removed from the beam In addition to this type of event protons (as well as all other charged particles of mass considerably larger than that of the electron) lose energy by imparting it to the electrons of the medium in which they move without being appreciably deflected from their trajectory the energy loss simply slows down the heavy particle Ultimately the heavy particle (proton) comes to rest and the distance it travels between the point where it has an energy E and the point where it comes to rest is called the range of the proton The range is a function of the energy of the proton and of the medium in which it moves A semiempirical formula connecting the energy and range in air (N T P) for protons is

$$R = \left(\frac{E}{9.3}\right)^{1.8}$$

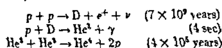
where R is in meters and E is in Mev This relation is valid between a few Mev and about 200 Mev For other substances one must consider the relative stopping power which is a number giving the reciprocal of the ratio of the thickness in g/cm² of two layers producing the same energy loss The stopping power decreases with Z the atomic number [E C S]

Proton proton chain

An energy releasing nuclear reaction chain which is believed to be of major importance in energy production in hydrogen rich stars The net effect of the proton proton chain is the conversion (fusion) of four protons into one He^4 (helium) nucleus with the release of about 26 Mev of energy *See* FUSION NUCLEAR NUCLEAR REACTION

The first (and slowest) reaction of the proton proton chain is the combination of two protons through a β decay process releasing a positron and a neutrino and forming a deuteron The deuteron then rapidly undergoes a p,γ reaction with another proton forming He^3 It is believed that the next reaction is a simple nuclear fusion reaction between He^3 nuclei to produce He^4 and release two protons which rejoin the chain

The estimated rates for these sequential reactions under typical stellar conditions are after E E Salpeter



$$p + p \rightarrow D + e^+ + \nu \quad (7 \times 10^9 \text{ years})$$
$$p + D \rightarrow He^3 + \gamma \quad (4 \text{ sec})$$
$$He^3 + He^3 \rightarrow He^4 + 2p \quad (4 \times 10^5 \text{ years})$$

The theoretical cross section for the p p reaction is far too small (about 10^{-47} cm² at 1 Mev) to permit it to be measured in the laboratory by any foreseeable technique Nevertheless the present state of theoretical understanding of β decay processes is such as to make a theoretical evaluation highly credible The other two reactions of the chain are easily measurable and are well documented experimentally

The ... that al ... the lab ... source ... of hyd ... in the sun is an example *See* STELLAR EVOLUTION [R F P]

Bibliography *See* CARBON NITROGEN CYCLE

Protophyta

A division of the plant kingdom comprising those forms of living organisms capable of living an independent existence as single cells There are a few exceptions in which the cells are arranged in intimate contact with one another in a chain The ...

... to mitotic or meiotic division can be detected *See* PLANT KINGDOM

Interdependence of the cells of the chain forming species has not been clearly demonstrated other than for the larger species of the blue-green alga *Oscillatoria* However, the rapid disintegration and loss of viability of the individual cells of the colorless sulfur bacterium *Beggiatoa* following mechanical injury, suggests that such interdependence may occur

The Protophyta differ from the single celled species in the Thallophyta in that the Protophyta ... (2) lack definite ...

and (4) rarely have cell walls composed of cellulose or of chitin (as in the yeasts) *See* BACTERIAL MOTILITY

... the single celled or ... ozoa—in ... nucleus, (2) have no contractile vacuoles (3) never ingest particulate food (4) ... ever, long ... and (5) ... and cilia of the Protozoa (and the flagellated Thallophyta classified by some taxonomists with the Protozoa) are composed of one central pair of fibrils surrounded by nine pairs of fibrils all within a common sheath *See* ANIMAL KINGDOM, PROTOZOA THALLOPHYTA

The Protophyta differ from all higher forms of the plant and animal kingdoms in their complete lack of organization of cells into tissues and specialized organs and in the almost complete absence of morphologically discernible organs for sexual reproduction

The division is divided into three classes—the Schizophyceae Schizomycetes and Microtatobiotes *See* MICROTATOBIOTES, SCHIZOMYCETES, SCHIZOPHYCEAE [V B D S]

Protoplasm

The comprehensive term for the substance which has the distinctive properties of life Although originated for the material of animal cells, the term *protoplasm* was soon broadened to apply to the similar contents of the cells of plants as well The importance of protoplasm as the site of all the physical and chemical processes of life was especially emphasized by Claude Bernard (1878), the founder of general physiology and it is in this field that the term *protoplasm* has always been in greatest favor

Among cytologists the term is generally applied to the entire living material of the cell the nucleus and cytoplasm combined and regarded as a unit However, with the recent acquisition of more detailed information on the properties of the various subdivisions of the cell there has been less and less reason to use the inclusive term *See* CELL PROTOPLAST

Physical properties An important part of the research on protoplasm has been concerned with its physical properties Protoplasm usually occurs as microscopic droplets, single uninucleate cells

However there are conspicuous exceptions of which the plasmodium of slime molds is a classic example (*see* MYXOMYCETES) Such plasmodia may contain thousands of nuclei without cell walls and have a measurable volume of several cubic centimeters Protoplasm is generally a translucent essentially colorless slimy substance resembling grossly the white of an egg Some protoplasms have the viscosity of a stiff gel others flow almost like water Protoplasm is typically 75% water but the variation in this respect is enormous and ranges from 99% in jellyfish to 15% or less in dried seeds Its specific gravity commonly lies bewteen 1 02 and 1 08 and it has a refractive index near 1 4 Studies of protoplasm with the polarizing microscope have shown it to have in some instances (notably the skeletal muscle fiber) a high degree of molecular organization *approaching a well ordered crystal* line state Protoplasm is cohesive and can be stretched to form thin strands It exhibits anomalous viscosity inversely proportional to the shearing force this has been compared with thixotropy of colloids and ascribed to a similar basis that is the presence of asymmetrical submicroscopic particles (*see* COLLOID) Indeed studies with the electron microscope have shown such particles in many types of protoplasm Commonly especially in plants there are watery spaces or vacuoles within the protoplasm which are separated from the remainder by a differentially permeable membrane It is also not rare to find flowing movements (cyclosis) within protoplasm which are detectable as migration of the formed bodies often as an intracellular circulation *See* CELL (BIOLOGICAL)

Protoplasmic similarity of organisms Although protoplasm from the most diverse forms of life is very similar in appearance and other physical properties early attempts failed to find a common structural pattern such as reticular fibrillar granular or emulsoid It now appears that most of these conditions are expressions of various types of cell specialization Likewise the early philosophical arguments as to which part of the protoplasm might be the exclusive repository of the properties of life are only of historical interest in the light of the demonstrations that specialized parts of cells such as the mitochondrial fraction exhibit in vitro a degree of physiological activity approaching that of the intact cell *See* CYTOPLASM MITOCHONDRIA CELL NUCLEUS *see also* CELL INCLUSIONS NONCYTOPLASMIC [A W P]

Mechanical properties Because protoplasm is the physical basis of life a knowledge of its mechanical properties is essential to understanding its nature Protoplasmic cohesion and physical factors affecting the cell surface such as elasticity and surface tension have been the subjects of much investigation This knowledge forms the basis of some of the concepts of cell division and cell movement The mechanical properties of protoplasm are not always easy or even possible to determine One must make observations on individual cells (rather than populations of cells as is possible in chemical determinations) Also it has been shown that the mechanical properties of protoplasm may vary under different conditions and it is important that the technique of measurement does not in itself bring about changes Fortunately certain types of cell lend themselves well to such studies and this discussion shall be largely concerned with those few cell types For a discussion of the physical concepts of importance to the following discussion *see* ELASTICITY SURFACE TENSION, VISCOSITY OF LIQUIDS

Protoplasmic cohesion Cohesion may be described as the force of attraction among the particles of a substance which tends to hold it together It is defined more specifically in terms of the tangential force exerted on a stationary plane by a moving parallel plane the space between the two *planes being filled with the substance* With centimeter gram second units of measurement the unit of viscosity is the poise Water has a viscosity of 0 01 poise or 1 centipoise

If the shearing force due to velocity of the moving parallel plane is increased the tangential force should increase proportionately This holds true if the material between the two planes has the same viscosity at different shearing forces that is if the material is a Newtonian fluid Some materials however do not show a constant viscosity at different shearing forces These are called non Newtonian fluids that is they have an anomalous viscosity Materials which show a decrease in viscosity with an increase in shearing force are thixotropic (for example gelatin) Those which show an increase in viscosity with increase in shearing force are dilatant for example suspensions of starch grains In general protoplasm shows either true Newtonian or thixotropic properties

The various methods of viscosity measurement cannot all be applied to the living cell For example a standard method involves determination of the rate at which a liquid flows through a tube of known dimensions under standard conditions The higher the viscosity the lower the flow rate (the viscosity is inversely proportional to the rate of flow) If this method were applied to protoplasm the resulting injury would invalidate the results

Brownian movement method This is a method of viscosity measurement which can be applied to living protoplasm Microscopic particles suspended in a liquid are found to undergo Brownian movement The extent of movement is inversely related to the viscosity of the suspension liquid and this is the basis for absolute measurements of viscosity The factors affecting Brownian movement are described in the equation

$$D_x^2 = \frac{RTt}{N3\pi\eta\alpha}$$

where D_x is the displacement of the particle along one axis R the gas constant T the absolute temperature t the time N the Avogadro number η the viscosity and α the radius of the particle

In practice the method involves determination of the linear rate of movement of a suspended particle Determinations can be tedious Individual protoplasmic particles can be followed or in some cases the movement of an entire layer of *granules* can be determined following centrifugation of the granules to one side of the cell Values of protoplasmic viscosity obtained with this method are in agreement with those determined by the centrifuge method described later Frequently differences in the extent of Brownian movement can be seen in different regions of the same cell Also, the same cells under different conditions, may show changes in the extent of Brownian movement Motion pictures of cells in tissue culture usually reveal a movement of cytoplasmic particles Death of the cells results in a cessation of Brownian movement, *an indication of* protoplasmic gelation

Falling sphere method Another method of measurement involves determination of the rate at which a sphere of known density and radius falls through a liquid under standard conditions The viscosity is inversely proportional to the rate of fall and the relationship is expressed in the equation

$$V = \frac{2g(\sigma - \rho)\alpha^2}{9\eta}$$

where V is the velocity of the falling sphere g the gravity constant σ the specific gravity of the sphere ρ the specific gravity of the medium α the radius of the sphere and η the viscosity

This method has been successfully applied to the living cell For example in certain plant cells the starch grains have been found to move under the influence of gravity when the cells are viewed from the side with a horizontal ...

am...

in t...

ech...

viec... of the nucleoplasm from the rate of fall of the nucleolus through the contents of the germinal vesicle In these cases relatively low values of protoplasmic viscosity have been obtained less than 8 centipoises for the *cells of a Vicia* (bean) *plant and 7 10* centipoises for the nucleoplasm

Centrifuge method Though the protoplasm may be of low viscosity cytoplasmic granules do not settle as a result of gravity alone If the d fferen ...

... cosity the ... the centrifugal force or the longer *the* centrifugation time required to show a given degree of stratification The *centrifugation* causes no apparent injury to the cells because development following centrifugation is normal enough to cause stratification of the cytoplasmic granules. This

technique has been very useful for the determination of absolute values of viscosity, *and in fact*, because of the ease with which it can be applied to living protoplasm it is perhaps the most widely used method For absolute measurements corrections must be made for the effects of the cell cortex and neighboring granules on the velocity of movement The granule free cytoplasm in the egg of the sea urchin, *Arbacia punctulata*, as determined by this method, is about 2 centipoises which is only twice the viscosity of water A value of 4 3 was ob...

varying *the shearing force* By changing the rate of speed of centrifugation the shearing force is changed and any non Newtonian properties become apparent The interior protoplasm of unfertilized *Arbacia* and *Cumingia* eggs is thus shown to have a constant viscosity whereas the cortex of the ameba appears to behave as a *thixotropic gel* The nucleoplasm of certain echinoderm eggs is also thixotropic *See* MICROSCOPE, CENTRIFUGE.

characteristic changes in cytoplasmic viscosity The appearance of the mitotic spindle is preceded by a sharp increase in cytoplasmic viscosity These alterations in viscosity reflect changes which are occurring at the molecular level and their elucidation will undoubtedly lead to a better understanding of the mechanism of cell division

It is also possible by means of the centrifuge method to detect changes in *the rigidity* of the cell cortex at different times during the process of cell division This is accomplished by determination of the centrifugal force required to dislodge the granules from the cell cortex

ing medium is due to the presence of a *definite* membrane at the cell surface It is probably *not correct* to speak of surface *tension* when referring to this membrane Measurements of tension at the cell surface are in fact, the result of a combination of surface and elastic tensions

Kinetic flow method The *kinetic flow method* clearly demonstrates the tension at the cell surface. Just before an *Arbacia* egg cell divides the two blastomeres *remain attached* by a stalk *If one* blastomere *is punctured* cytoplasm flows from the undamaged blastomere through the stalk because of the excess internal pressure Measurements of the resultant decrease in volume are obtained *from* motion pictures of the process From this the rate *of flow of* cytoplasm through the *stalk is* obtained *The excess internal* pressure is calculated according to Poiseuille's law for the flow of liquid *through* a tube using figures for the dimension of the stalk the viscosity of the cytoplasm and the rate of flow The internal pressure is measured at

64 dyne/cm², and the tension at the cell surface, 0 09 dyne/cm

Compression method Perhaps the most accurate method for measurement of tension at the cell surface is the compression method This has been applied to the spherical egg cells of *Arbacia* A microbeam of gold 6 μ thick and 180 μ wide is used to compress the cell The microbeam is always held parallel to the plane of compression The instrument is calibrated so that the deflection of the microbeam can be related to the absolute force involved The following equation is used in calculating the tension

$$F/A = P = T(1/r_1 + 1/r_2)$$

in which F is the force necessary to flatten the cell to a given extent, A is the area of the flattened portion of the cell, T, the surface force, r, the radius of curvature of the unflattened bulging portion of the cell as determined from a side view photograph of the compressed cell (r_1 represents one side of the cell and r_2 the other side) The actual measurements are obtained from photographs of the flattened cells When the cells are compressed 25 μ the tension at the cell surface is 0 133 dyne/cm When the cells are compressed to a smaller extent, the tension is less This relationship between compression and tension is the result of the elastic properties of the cell surface Tension is plotted as a function of compression If the resulting curve is extrapolated to zero compression, a value of 0 08 dyne/cm for the uncompressed egg is obtained

Centrifuge method The centrifuge method is a technique which is particularly useful for comparative studies of cell surface tension The cells are placed in a medium with a density close to that of the cells and centrifuged in a microscope-centrifuge During centrifugation there is a separation of the various intracellular components The oil droplets migrate to the light half, and the yolk to the heavy half In a medium of appropriate density, and with sufficient centrifugal force the light and heavy halves of the egg are pulled apart The centrifugal force which is just sufficient to separate the two halves regardless of the centrifugation time is the value used to calculate the tension

$$\tau DT = C_g[V_h(\rho_h - \rho_m) + V_L(\rho_m - \rho_L)]$$

where T is the surface tension, D the diameter of the cylinder (the diameter of the drawn out cell at the moment of instability), C_g the centrifugal force, V_h, the volume of the heavy half, V_L, the volume of the light half, ρ_h, the density of the heavy half, ρ_m, the density of the medium and ρ_L, the density of the light half The method has been applied to a variety of cells Rabbit macrophages showed a value of 2 dyne/cm, frog leukocytes 1 3, *Amoeba dubia*, 1–3 This method is best adapted for comparative studies rather than for the determination of absolute values For example changes in tension at the cell surface at the time of fertilization or following treatment with various salt solutions or during the aging of cells, are appropriately studied with this method

Other methods Other methods of measuring tension at the cell surface have been applied with some success The sucking method involves sucking a portion of the cell surface into a capillary tube The negative pressure required to deflect the cell surface to a given degree can be determined Deflection of the surface is directly proportional to the negative pressure The mathematics of this method have not been completely worked out

The sessile drop method has been applied to a small number of cells The surface tension of a drop of liquid can be determined from the form of a flattened drop Some cells which presumably have surface membranes with low rigidity, have been measured in this way Many cells, however, do not show any flattening under the influence of gravity

The stretching method involves determination of the force required to stretch a cell to a given degree A micromanipulator is used to stretch the cell, and one of the needles is calibrated so that the force required to cause a given deflection of the needle is known The values for tension determined by this method agree quite well with those obtained with the compression method *See* MICROMANIPULATION

Data Most of the methods for measurement of tension at the cell surface indicate relatively low values Determinations by the compression method for example, give a value of only 0 08 dyne/cm for the undistorted cell The surface tension at an air water interphase on the other hand is about 73 dyne/cm E N Harvey has pointed out that the tension of cells is at least 1/1000 of this value This is of decided advantage to a cell with its high ratio of surface to volume If the tension were much higher, it would require the expenditure of much more energy to move or to adjust its shape to neighboring cells This then as Harvey points out represents an adjustment of mobile cells to physical constraints imposed by the environment *See* BIOPHYSICS, CELL (BIOLOGICAL), CELL DIVISION

[C V H]

Bibliography K S Cole Surface forces of the *Arbacia* egg *J Cellular Comp Physiol* 1(1) 1–8 1932 L V Heilbrunn *An Outline of General Physiology* 3d ed 1952, L V Heilbrunn (ed) *Handbuch der Protoplasmaforschung* vol 2 1954 S L Palay (ed) *Frontiers in Cytology* 1958 W Seifriz *Protoplasm* 1936 E B Wilson *The Cell in Development and Heredity*, 3d ed 1925 W L Wilson Rigidity of cell cortex during cell division *J Cellular Comp Physiol*, 38 409 415 1951

Protorosauria

An order of Permian and Triassic reptiles (subclass Euryapsida) whose systematic affinities are under dispute Most of the forms included in this group are incompletely or poorly preserved The best known and most adequately studied genera are *Araeoscelis* *Trilophosaurus* *Macrocnemus* and *Tanystropheus* The last two genera [illegible] ther with

Tanystropheus longobardicus after a specimen 4 3 meters in length Triassic Switzerland (From B Peyer 1944)

Askeptosaurus regarded by some authorities as a thalattosaurian) from the Alpine Middle Triassic have also been interpreted as closely related derivatives of diapsid perhaps eosuchian stock

One of the striking aspects of the skeleton of these forms is the elongation of the neck region without notable increase in the number of vertebrae this reaches absurd dimensions in *Tanystropheus* where the neck is as long as the body and tail combined Generally the forelegs are much shorter than the hindlegs and the digits terminate in horny claws These features tend to indicate terrestrial habits (even though some forms are known only from marine deposits) Any discussion of phylogenetic relationships among the members

[R Z]

Bibliography J Piveteau (ed) *Traite de Paleontologie* vol 5 1955 A S Romer *Osteology of the Reptiles* 1956

Prototheria

One of the four subclasses of the class *Mammalia* Prototheria contains a single order the Monotremata No ancestral genera of fossil monotremes are known and the structure of the living monotremes is so specialized that the affinities of the Prototheria are largely conjectural Most mammalogists believe that the prototheres arose from a different stock of therapsid reptiles than the one that gave rise to the Theria This would mean that the history of the monotremes has been separate from the history of other mammals for at least 175 000 000 years

No fossils earlier than the Pleistocene are known and these come from Australia The duck billed platypus *Ornithorhynchus anatinus* and several species of the spiny anteater *Tachyglossus* are living representatives of this group They are found in the Australian region Everything indicates that the Prototheria represent a very small and relatively unsuccessful group that has miraculously survived in an isolated corner of the earth *See* MAMMALIA MONOTREMATA *see also* THERAPSIDA

[D D D.]

Prototype (equipment)

The term used at that stage in the design process when the component has been realized physically in a form that satisfies the functional environmental reliability maintainability packaging and other requirements but when the design does not necessarily reflect the techniques of manufacture by which it will be ultimately produced in quantity (see SYSTEMS ENGINEERING) Although considerations of the methods of manufacture play a role in the preliminary stages of the evolution of a component early developmental versions since they are produced in small quantities (one to several units) may not be designed to take advantage of the methods of fabrication peculiar to large scale production

The prototype equipment is manufactured on a relatively small scale which might not justify for instance, expensive fixtures and dies or molds, which would be necessary in subsequent large-scale manufacture. Those parts which are especially suited to large-scale manufacture however are reproduced as closely as possible to the form they will take when they are made in large scale manufacture. It is at this stage that the manufacturing process is most seriously considered Several prototypes embodying different manufacturing techniques in several of their elements might be explored before the final design is decided upon For example the main structural element might be designed as a casting or a weldment or fabricated out of metal stampings The final manufacturing prototype design will embody most of the ideas with regard to manufacturing techniques that are desired in the final product *See* PILOT PRODUCTION PRODUCTION ENGINEERING

The final prototype as well as some of the preliminary ones is subjected to the same scrutiny as production units (*see* QUALIFICATION TEST) in order to demonstrate the satisfactory performance of the component under all operating circumstances

[R W M]

Protozoa

A phylum of primitive animals the majority of which are microscopic They vary in length from 2 or 3 microns (μ) to several inches In practice the phylum also includes certain groups such as the slime molds and phytoflagellates which botanists consider to be plants Free living species occur in the soil and in fresh salt and brackish waters Endoparasites like the malarial parasites trypanosomes and gregarines are located in the body cavities or tissues of higher organisms. Many species are uninucleate others are binucleate or multinucleate. Although some species form colonies Protozoa lack tissues and organs

TAXONOMY

Classification of the Protozoa is not yet stabilized and, in that sense, the system outlined below must be considered tentative. Certain changes suggested by J. O Corliss for the ciliates are included

Phylum Protozoa
- Subphylum 1 Mastigophora
  - Class 1 Phytomastigophorea
    - Order 1. Chrysomonadida
    - Order 2 Heterochlorida
    - Order 3 Cryptomonadida
    - Order 4 Dinoflagellida
    - Order 5 Phytomonadida
    - Order 6 Euglenida
    - Order 7 Chloromonadida
  - Class 2 Zoomastigophorea
    - Order 1 Rhizomastigida
    - Order 2 Protomastigida
    - Order 3 Polymastigida
    - Order 4 Trichomonadida
    - Order 5 Hypermastigida
- Subphylum 2 Sarcodina
  - Class 1. Actinopodea
    - Order 1 Helioflagellida
    - Order 2 Heliozoida
    - Order 3 Radiolarida
  - Class 2 Rhizopodea
    - Order 1 Proteomyxida
    - Order 2 Mycetozoida
    - Order 3 Amoebida
    - Order 4 Testacida
    - Order 5 Foraminiferida
- Subphylum 3 Sporozoa
  - Class 1 Telosporidea
    - Subclass 1 Gregarinida
      - Order 1 Eugregarinida
      - Order 2 Schizogregarinida
    - Subclass 2 Coccidia
    - Subclass 3 Haemosporidia
  - Class 2 Cnidosporidea
    - Order 1 Myxosporidia
    - Order 2 Actinomyxida
    - Order 3 Microsporida
    - Order 4 Helicosporida

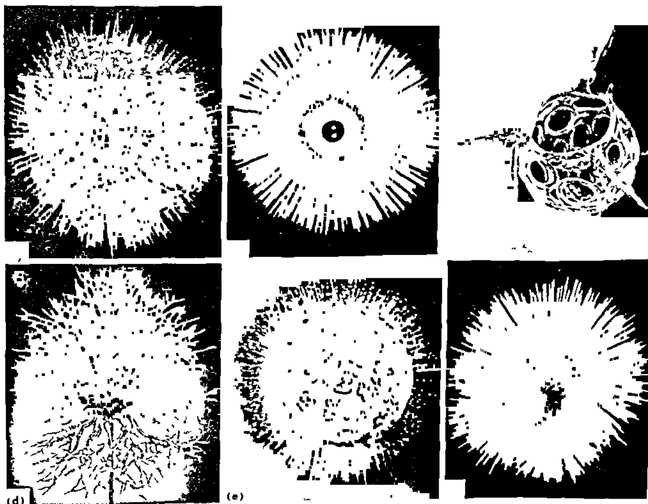


Fig 1 Glass models of marine protozoans (a) *Trypanosphaera transformata* Haeckel (Indian Ocean) (b) *Actissa princeps* Haeckel (Indian and Pacific Oceans) (c) *Peridium spinipes* Haeckel (Pacific Ocean) (d) *Lithocircus magnificus* Haeckel (Atlantic Ocean) (e) *Collozoum serpentinum* Haeckel (Atlantic Ocean) (f) *Globigerina bulloides* d Orbigny (all seas) (*The American Museum of Natural History*)

Class 3 Acnidosporidea
Subclass 1 Sarcosporidia
Subclass 2 Haplosporidia
ubphylum 4 Ciliophora
Class 1 Ciliata
Subclass 1 Holotricha
Order 1 Gymnostomatida
Order 2 Suctorida
Order 3 Chonotrichida
Order 4 Trichostomatida
Order 5 Hymenostomatida
Order 6 Astomatida
Order 7 Apostomatida
Order 8 Thigmotrichida
Order 9 Peritrichida
Subclass 2 Spirotricha
Order 1 Heterotrichida
Order 2 Oligotrichida
Order 3 Tintinnida
Order 4 Entodiniomorphida
Order 5 Ctenostomatida
Order 6 Hypotrichida

MORPHOLOGY

The protozoan body may be plastic as in ameboid species In many others changes in form are limited by a moderately flexible to rather rigid

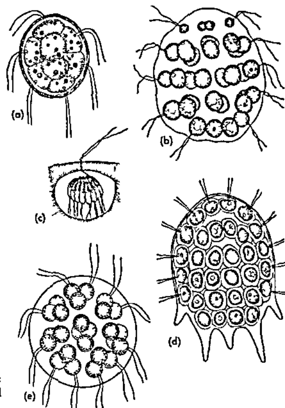


Fig 3 Spheroid colonies of Phytomonadida (a) *Pandorina* (b) *Pleodorina illinoisensis* (after Merton) (c) Sperm packet *Pleodorina* (after Merton) (d) *Platydorina caudata* (after Kofoid) (e) *Eudorina* (L H Hyman The Invertebrates vol 1 McGraw Hill 1940)

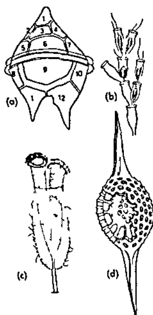


Fig 2 Types of protozoan encasements. (a) Theca of the dinoflagellate *Peridinium* showing separate plates (after Barrows) (b) Lorica of *Dinobryon* one of the Chrysomonadida (after Kent) (c) *Cothurnia* (order Peritrichida) two zooids within a lorica (d) A radiolarian skeleton of the siliceous type (after Haeckel) (From L H Hyman The Invertebrates vol 1 McGraw Hill 1940)

tions however are common Bilateral symmetry is rare since most active swimmers show more or less pronounced spiral torsion Armored species (Fig 2a–d) possess an outer test (or shell) or theca which often contains cellulose and is analogous to the cell wall in higher plants The theca of many dinoflagellates is made up of plates arranged in specific patterns Tests are composed mostly of inorganic material as in Foraminiferida and Testacida Other skeletal structures include complex radiolarian skeletons (Fig 2*d*)

Colonies and aggregates In colonial species the zooids are joined together in some characteristic pattern In spheroid colonies (Fig 3) a matrix is secreted by the associated organisms during growth of the colony In arboroid colonies (Fig 4*h k*) the zooids may be attached to a branching stalk or live in loricae attached to one another

val stages In some Volvocidae reproductive and vegetative zooids are distinguishable More often colony members are morphologically similar

In addition to colonies Protozoa of certain species may form aggregates by repeated fission without prompt separation of daughter organisms Palmella stages of phytomonad flagellates analogous to spheroid colonies are such aggregates of non

flagellated individuals The chains of certain dino flagellates like *Gonyaulax catanella* represent other aggregates

Locomotor organelles Locomotor organelles include pseudopodia flagella and cilia The several kinds of pseudopodia are retractable extensions of the body Axopodia are slender usually with an axoneme and radiate (usually singly) from the body surface Filopodia also are slender hyaline and taper from base to tip but they have no axoneme and tend to branch and anastomose Lobopodia are relatively broad and have rounded tips, the larger lobopodia show a granular endoplasm Myxopodia or rhizopodia are filamentous

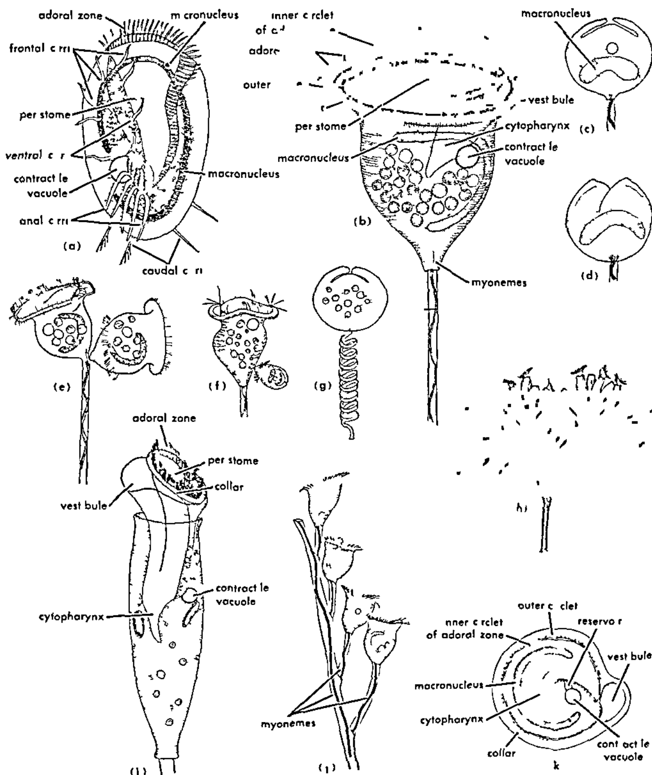


Fig 4 Hypotricha and Peritricha (a) *Euplotes* (b) *Vorticella* (c d e) Stages of fission in *Vorticella* (f) Macro- and microconjugants in *Vorticella* (after Kent) (g) *Vorticella* stalk contracted (h) Arboroid colony *Opercularia* (i) Zooid from *Opercularia* colony (j) Portion of colony (*Carchesium*) showing stalk muscle (after Conn) (k) Peristome of *Vorticella* seen from above (after Noland and Finley) (From L H Hyman The Invertebrates vol 1 McGraw-Hill 1940)

d branch and anastomose to form networks
xopodia and myxopodia have a sticky outer layer
which a flow of cytoplasmic granules is charac
ristic In the other types of pseudopodia the
uter layer is denser than the inner cytoplasm

Flagella Flagella found in Mastigophora and
n flagellated stages of Sarcodina and Sporozoa
how an outer sheath enclosing a matrix in which
in axoneme extends from a self reproducing gran
ule the blepharoplast in the cytoplasm In elec
tron micrographs the axoneme is a bundle of fi
brils 2 central and 9 peripheral in certain spe
cies Delicate filaments mastigonemes" or "Flim
mer" project from the sheath in certain species
The number of flagella ranges from one or two in

along the body or may form the margin of an un
dulating membrane In addition to flagellar ax
onemes such organelles as the axostyle parabasal

volves the principle of the screw The basic func-
tion of the flagellum would generally be to cause
rotation of the flagellate on its major axis and gy

h typically
ecture even
From the

basal granule of each cilium a filament extends
cytoplasm to join others in forming a
desma
the
rdi

neuromotor apparatus
nates locomotor activities

Groups of cilia may be combined to form com
pound ciliary organelles Membranes represent
fused longitudinal rows of cilia membranelles
several fused transverse rows (Fig 4a) Cirri rep-
resent fused tufts of cilia 3–20 or more Mem
branes and membranelles usually occur in a buccal
cavity leading to the cytostome Cirri are primarily
locomotor organelles

Contractile myonemes occur typically in the cor
tex of ciliates such as *Spirostomum* and *Stentor*,
which can change shape rapidly Similar myone
mes are found in many gregarines Analogous con
tractile fibrils occur in the stalks of certain ciliates
like *Vorticella* (Fig 4b g k), *Carchesium* and
Zoothamnium

Trichocysts Trichocysts are found in the outer
cytoplasm of certain ciliates and flagellates Mu
coid trichocysts are elongated bodies which under
artificial stimulation may be ejected without ap-
preciable change in form In situ they may be
vitally stained with neutral red and other dyes Fil
amentous trichocysts as seen in *Paramecium* *Fron
tonia* and others show upon discharge a pointed
tip and a cross striated shaft In electron micro-
graphs the shaft shows transverse striations with

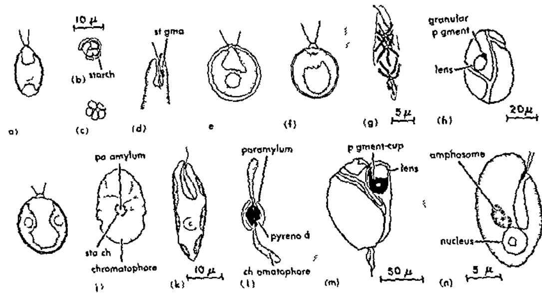


Fig 5 Chromatophores ocellus amphosome and pyrenoids (a) Chromatophores in *Chlamydomonas agloeformis* (after Pascher) (b c) Compound pyrenoids of *Pyramidomonas* with adherent starch (after Geitler) (d) Stigma of *Euglena* (e) Chromatophores in *C umbonata* (after Pascher) (f) Chromatophores in *C inversa* (after Pascher) (g) Chromatophores in *Euglena geniculata* (after Hollande) (h) Ocellus in *Pouchetia* (after Kofoid and Swezy) (i) Chromatophores in *C biococca* (after Pascher) (j) Chromatophore in *Peridinium umbonatum* (after Geitler) (k) Chromatophores in *Colacium* (after Johnson) (l) Chromatophore with pyrenoid and paramylum (after Hollande) (m) Ocellus in *Erythropsis* (after Kofoid and Swezy) (n) Amphosome (stained) in *Cryptomonas* (after Hollande) (From R P Hall Protozoology Prentice-Hall 1953)

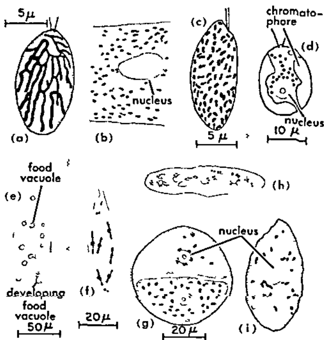


Fig 6 Mitochondria, granules and osmophilic inclusions (a) Mitochondria in *Polytoma* (after Hollande) (b) in *Protoopalina* (after Richardson and Horning), (c) in *Chilomonas* (after Hollande) (d) Granules in *Chlamydomonas* (after Dangeard), (e) in *Paramecium* (after Dunihue), (f) in *Euglena* (after Dangeard) (g) Osmiophilic inclusions in *Gregarina* (gametocytes in cyst) (after Joyet Lavergne), (h) in *Paramecium* (after Dunihue), (i) in *Protoopalina* (after Richardson and Horning) (From R P Hall, *Protozoology*, Prentice Hall, 1953)

the same periodicity (about 55 mμ) in various holotrichous ciliates In *Oxyrrhis*, supposedly a primitive dinoflagellate discharged trichocysts show a similar striated shaft in which the periodicity is about 60 mμ So called toxicysts as found in *Dileptus*, after discharge show a nonstriated tubular portion attached to a basal capsule and a rodlike tip Toxicysts upon contact with other microorganisms, produce paralysis or cytolysis in susceptible species The trichocysts of *Paramecium* can be partially extruded to anchor the ciliate during active feeding Mucoid trichocysts, especially in phytoflagellates, may represent material accumulated for cyst walls and comparable layers

Inclusions. Contents of the endoplasm may include a nucleus or nuclei, food vacuoles in holozoic species chromatophores and stigma in many phytoflagellates, mitochondria and osmiophilic inclusions stored food reserves and such organelles as axostyles and parabasal apparatus in particular flagellates

Chromatophores Many phytoflagellates have chromatophores (Fig 5) These contain chlorophyll although the green color may be masked by other pigments (such as red, blue brown and yellow) Specific types of pigments show some correlation with taxonomy In electron micrographs, chromatophores show a membrane enclosing thin opaque lamellae which range from 18-32 mμ, and alternate with slightly thicker layers of transparent material In certain strains of *Euglena*, chromatophores have been eliminated by exposing the flagellates to streptomycin high temperature (36°C), and Pyribenzamine Pyrenoids (Fig 5*b,c,l*) are often associated with chromatophores as in the Phytomonadida and certain Euglenida, but their taxonomic distribution may be erratic In Euglenida for example, pyrenoids may be present or absent in different species of one genus Pyrenoids range from homogeneous bodies to clusters of granules and are embedded in or closely applied to chromatophores Pyrenoids are usually assumed to participate in synthesis of polysaccharides

Stigma A stigma (Fig 5*d*) occurs in many green flagellates and some colorless strains Studies on behavior, including comparisons of stigma free mutants with normal *Chlamydomonas*, show that the stigma is a photoreceptor involved in responses to light The usual stigma contains reddish pigments, a plate of granules in *Euglena*, presumably carotenoids The stigma often persists in bleached *Euglena* grown in light An ocellus a more complex photoreceptor showing a lenslike body enclosed in pigment, occurs in certain dinoflagellates (Fig 5*h,m*)

Mitochondria Mitochondria (Fig 6*a b,c*) are more or less elongated inclusions sometimes filaments, which in electron micrographs show a double wall a matrix, and either folds (cristae) or fingerlike microvilli extending into the matrix Cristae have been reported in *Euglena* and certain other phytoflagellates, microvilli in *Amoeba* and several ciliates In addition to mitochondria, small globules which can be vitally stained with neutral red are often present (Fig 6*d e f*) Many species also show osmiophilic inclusions of comparable size Some osmiophilic inclusions (Fig 6*g i*) have been considered protozoan Golgi material

Food reserves Food reserves are often stored as discrete bodies in the cytoplasm Polysaccharide reserves include paramylum a glucose polymer of Euglenida, leucosin of Chrysomonadida, starches of other phytoflagellate orders glycogens of ciliates, amebas various flagellates and Sporozoa Under certain conditions many species store lipids as small or large globules Crystals as seen in *Amoeba*, may represent stored food in some cases

Contractile vacuoles Contractile vacuoles (Fig 7*b f,g*) occur in most fresh water Protozoa and in certain parasitic and marine types Such a vacuole shows a rhythmic cycle that is an origin growth in volume and discharge of contents The growing vacuole may receive fluid from one or more contributory canals as in *Paramecium* from fusion with small cytoplasmic vacuoles as seen in *Euplotes*, or perhaps by secretion through the vacuolar membrane as occurs in *Eudiplodinium* Fluid is discharged through a pore which may be temporary as in *Amoeba* or may show a specific position as in the ciliates Periodicity of the cycle varies with temperature salinity, pH of the medium and with activity of the organism A presum

ably analogous contractile tube opening externally through several pores occurs in the parasitic cili ate *Haptophrya* (Fig 7a) The major function of the contractile vacuole is that of eliminating ex cess water which reaches the cytoplasm through endosmosis with ingested food or by arising in metabolism Waste products may be eliminated in the vacuolar fluid but the relative importance of this function is uncertain In addition to contrac tile vacuoles so-called sensory vacuoles occur in certain parasitic and free living ciliates (Fig. 7c–e)

Nuclei Ciliates have two kinds of nuclei the mi cronucleus and macronucleus (Fig 4a) and some times more than one of either or both The macro-nucleus may be a polyploid nucleus produced in at least certain species by repeated division of chro-mosomes in the macronuclear anlage arising in conjugation or autogamy Protozoa other than cili ates have nuclei of one kind although often more than one nucleus In some species structure of the nucleus may change during growth It is reported that in certain radiolarian cycles the nucleus of the young organism develops into a polyploid nu cleus of the adult Protozoan nuclei in addition to chromosomes usually contain one or more non chromatinic bodies such as the endosome or nu cleolus Except for the macronucleus of ciliates, nuclear division involves mitosis

NUTRITION

Food in solution may pass through the body wall in saprozoic feeding while solid particles pass through a temporary or permanent cytostome in holozoic feeding Herbivores feed mainly on bacte-ria or algae carnivores on other Protozoa or small Metazoa omnivorous types on a variety of micro-organisms

Holozoic nutrition Holozoic Protozoa take their food into food vacuoles formed at the base of a gullet or tentacle as in the Suctorida or in ameboid types the vacuole is derived from the surface of the body during ingestion After ingestion con tents of the vacuole soon become acid Digestion follows and the products are absorbed During this period there is typically a rise in pH of the vac uolar contents Undigested materials are eventually eliminated In ciliates this typically occurs through a particular area (cytopyge) in the body wall

Adult Suctorida typically have tentacles which function in feeding The capitate tentacle shows a terminal bulb (possibly a tuft of papillae em bedded in amorphous material) an outer sheath and an inner tube extending into the endoplasm In feeding a tentacle adheres to a captured ciliate Protoplasm soon begins to flow down the inner tube into an enlargement which becomes a food vacu ole The specific mechanism involved in the flow whether it be suction peristalsis or positive pres sure from the punctured body of the prey is still uncertain

Nutritional requirements Food requirements include minerals sources of nitrogen and carbon and usually one or more vitamins Minimal re quirements have been determined for some species in axenic cultures An ammonium salt or a nitrate is adequate as the nitrogen source for many phy toflagellates Other Protozoa need one or more amino acids For example *Tetrahymena pyri formis* needs 10 or 11 amino acids depending upon the strain These are arginine histidine isoleucine leucine lysine methionine phenylalanine threo-nine tryptophan valine, and usually serine The remaining amino acids of proteins are synthesized

pounds such as ethanol acetate butyrate lactate, pyruvate glucose and others in various species as a source of energy and material for synthesis Suit ability of particular compounds varies with the species For example glucose is excellent for cili ates and many other Protozoa but not for *Euglena*

Trace minerals reportedly needed by one or more species include calcium cobalt, copper iron potassium magnesium manganese phosphorus sulfur and zinc. Also growth of certain species may be stimulated by aluminum boron barium io-dine sodium silicon or vanadium

Vitamin requirements range from none in certain phytoflagellates such as some species of *Brachio*

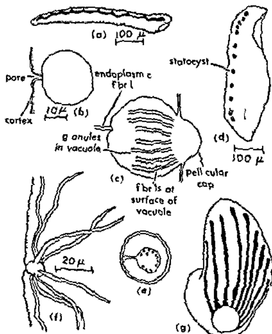


Fig 7 Contractile vacuoles and tubes (a) Contrac tile tube in *Haptophrya* an intestinal ciliate from a salamander (after MacLennan (b) Contractile vacuole and pore in *Eudiplodinium* a parasitic ciliate from ruminants (after MacLennan) (c d e) Sensory vacuoles of *Blepharoprosthium* (after Dogiel) and *Loxodes* (after Penard) (f g) Contractile vacuole and con tributory canals in *Paramecium* (after King) and *Tillina* (after Turner) (From R P Hall Protozoology Prentice-Hall, 1953)

monas Chlorogonium and *Chlamydomonas*, to six or more in typical heterotrophs Such differences in requirements indicate differences in the extent to which particular species can synthesize vitamins utilized in metabolism Thiamin and vitamin B_{12} are most commonly required by phytoflagellates In a few phytoflagellates an additional need for biotin or rarely, for riboflavin (in *Peranema*) has been reported Additional vitamins required by various other Protozoa include pyridoxine pantothenic acid nicotinic acid (or nicotinamide), folic acid, and thioctic acid Additional requirements occasionally include a sterol and one or more of the purine and pyrimidine components of nucleic acids The presence of chromatophores does not eliminate the need for vitamins in many phytoflagellates

Knowledge of specific vitamin requirements has been applied to microbiological assays involving Protozoa such as thioctic acid in *Tetrahymena pyriformis* and vitamin B_{12} in *Ochromonas malhamensis* and *Euglena gracilis*

REPRODUCTION

Reproduction in the less specialized Protozoa involves binary fission (Fig 4*c–e*) or simple budding Budding may be external or internal Internal budding seen in certain Suctorida, involves formation of a bud within a brood pouch an almost completely closed invagination in the parental body After the bud is formed, it is released through a "birth pore" In external budding the bud is pinched off at the surface of the parental body Multinucleate types often reproduce by schizogony (Fig 8*a b*) or, in some cases by plasmotomy In schizogony a number of buds become separated from a residual mass of cytoplasm In plasmotomy (Fig 8*c d*) a multinucleate organism divides into several cells, each with a number of nuclei As a part of reproduction, nuclei divide either before or during division of the body Other organelles vary in behavior Parental flagella may be retained or, in certain species resorbed at the beginning of reproduction In either case, one or more new flagella must be produced to equip the daughter organisms Blepharoplasts of flagella and basal granules of cilia seem to be self reproducing So is the kinetoplast (parabasal body) of trypanosomes Various other parental organelles, such as the axostyles and typical parabasal bodies as a rule, are more or less completely resorbed and new ones are produced in the daughter organisms Fission of ciliates may involve extensive reorganization of the body There may be a resorption of locomotor organelles in certain groups and the formation of a new mouth with its associated organelles

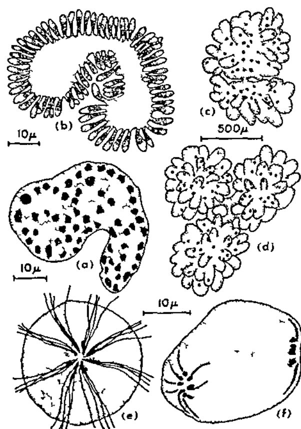


Fig 8 Reproduction (a) Schizont of *Ovivora thalassemae* (after Mackinnon and Ray) (b) Schizogony in *O thalassemae* (after Mackinnon and Ray) (c, d) Plasmotomy in *Pelomyxa carolinensis* (after Kudo) (e) *Coronympha octonaria*, vegetative stage showing nuclei and flagellar groups (after Kirby) (f) Nuclear groups at end of telophase, just before plasmotomy in *C octonaria* (after Kirby) (From R P Hall, *Protozoology*, Prentice-Hall, 1953)

Life cycles In simple cases life cycles may include merely active and encysted stages Specialization may involve dimorphism or polymorphism in the active phase (or sometimes in encysted stages) or the addition of a sexual phase, which may or may not be obligate Dimorphism is represented by (1) larval and adult stages in Suctorida (2) flagellate and ameboid stages in certain Mastigophora and Sarcodina (3) flagellate and nonflagellated stages like the various phytomonad flagellates which form palmella stages (4) amebic and plasmodial stages as seen in the Mycetozoa

Encystment involves secretion of one or more membranes to form a cyst wall Depending upon the species the wall may or may not contain one or more emergence pores closed by thin membranes Precystic activities include accumulation of reserve food resorption of locomotor organelles changes in form (toward the spherical in many species) and loss of water Protective cysts (Fig 9*a–d*) have fairly thick walls and in some species may be resistant to desiccation For example dried cysts of *Colpoda cucullus* have remained viable for about 5 years Within reproductive cysts (Fig 9*e f*) fission or budding or sometimes gametogenesis and syngamy, occur in various species

Excystment involves absorption of water and rupture of cyst membranes along with necessary reorganization of the body such as the regeneration of locomotor and feeding organelles in many species

Sexual reproduction Sexual activities include syngamy, conjugation and in some ciliates autogamy These processes involve meiosis

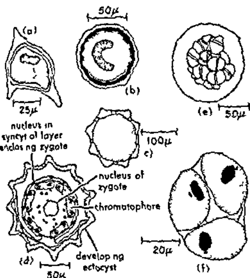


Fig 9 Protective and reproductive cysts (a) Protective cyst of Ceratium (after Beers) (b) of Didinium (after Beers) (c) of Bursaria (after Beers) (d) Encysted zygote of Volvox (after Janet) (e) Reproductive cyst of Gyrodinium a dinoflagellate (after Kofoid and Swezy) (f) of Colpoda a ciliate (after Kidder and Claff) (From R P Hall Protozoology Prentice Hall, 1953)

by the fusion of haploid gametic nuclei to produce a diploid zygotic nucleus In life cycles of the phytomonad flagellates certain Coccidia and gregarines meiosis occurs in an early division of the zygote and the organisms are haploid except in the zygote stage In other cases including conjugation and autogamy meiosis occurs prior to the formation of gametic nuclei as a result the organisms are diploid except in the gamete stage

In syngamy [illegible]

[illegible] as in malarial parasites Voltox and typical Coccidia In conjugation the two conjugants are usually similar In exceptional cases like *Vorticella* and relatives conjugation involves a microconjugant and a macroconjugant (Fig 10*e f*) only the latter survives.

Conjugation involves pairing of ciliates (Fig 10) followed by divisions usually *three of the micronucleus* Reduction of chromosomes to the haploid number usually occurs in the first division *Some of the daughter* nuclei ordinarily degenerate leaving two haploid gametic nuclei in each conjugant A *migratory* gametic nucleus is transferred from each conjugant to its mate *where it joins the* stationary gametic nucleus to form a zygotic nucleus This diploid nucleus now divides one or *more times* depending upon the species, and *the products* differentiate into micronuclei and macronuclei The old macronuclei degenerate during conjugation

In autogamy which occurs typically *but not exclusively* in unpaired ciliates nuclear *behavior is similar* except that there is no exchange of gametic nuclei Instead the gametic nuclei fuse to form a zygotic nucleus in the ciliate which produced them

In order to conjugate ciliates must belong to [illegible]

and each variety contains two or more mating types In general conjugation occurs readily between mating types belonging to a single variety, but much less readily or not at all between mating types belonging to different varieties Also the viability of survivors from intervarietal conjugations is usually low In addition to the necessity for appropriate mating types conjugation is favored by starvation and in at least certain species, by the time elapsed since the last conjugation (maturity factor) Environmental conditions, such as temperature must also be favorable

Syngamy, in isogamous phytomonad flagellates such as *Chlamydomonas*, apparently requires two types of gametes and is favored by nitrogen-starvation

According to the concept of a "physiological life cycle" strains of ciliates pass through *a sequence* of phases under laboratory conditions (1) youth, (2) a phase of maturity in which conjugation can occur (3) old age in which the ciliates undergo senescence and conjugation becomes impossible Supposedly the occurrence of senescence *can* be prevented by conjugation or by autogamy in certain species Although it is *now clear* that some

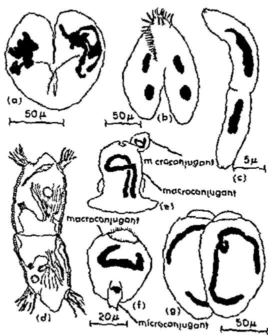


Fig 10 Conjugation (a) Nyctotherus (after Wichterman) (b) Pleurotricha (after Manwell) (c) Ancistrocoma (after Kofoid and Bush) (d) Cyclopposthium (after Dogiel) (e) Scyphidia (after Thompson, Kirkegaard and John) (f) Vorticella (after Finley) (g) *Euplotes* (after Turner) (From R P Hall Protozoology, Prentice Hall 1953)

species can be maintained indefinitely without conjugation, there is a possibility that autogamy replaces conjugation in some of these cases However, this cannot be true for the amicronucleate strains of *Tetrahymena pyriformis*, some of which have been maintained for more than 30 years

Parasitic species. Parasitic species are found in the four major divisions of Protozoa All Sporozoa are parasitic, as are several orders of flagellates (Trichomonadida, Hypermastigida) and ciliates (Astomatida, Apostomatida, Entodiniomorphida, Thigmotrichida) In addition, parasites occur in smaller taxonomic groups of Mastigophora, Sarcodina, and Ciliophora Protozoa also serve as hosts for protozoan parasites like the Suctorida, small amebas, and Sporozoa as well as for bacteria, fungi, and algae Some of the algal parasites like *Chlorella* in *Paramecium bursaria* are considered symbiotes

Many protozoan parasites (commensals) are more or less harmless Others like certain intestinal flagellates of termites and cellulose digesting ciliates of ruminants are considered symbiotic, in that they are beneficial to their hosts A relatively few species are distinctly pathogenic Such pathogens include organisms causing amebiasis kala azar, African sleeping sickness, Chagas' disease, malaria tick fever of cattle and other diseases *See* AMEBIASIS, CHAGAS' DISEASE, LEISHMANIASIS, MALARIA, SLEEPING SICKNESS, AFRICAN

Harmful effects may be produced in various ways Individual cells may be invaded and destroyed in malaria leishmaniasis and Chagas' disease, or tissues may be destroyed as in abscesses and ulcers involving *Entamoeba histolytica* The production of specific exotoxins by parasitic Protozoa is doubtful, although potent toxins are produced by certain free-living flagellates such as *Prymnesium* and *Gonyaulax*

Protozoan parasites may be transferred by such vectors as insects, ticks, by contamination of food or water, or by bodily contact In some cases parasites may pass through the placenta from mother to fetus (as in trypanosomiasis and occasionally malaria) Or, as in *Babesia bigemina* parasites invade eggs in the ovary of a vector (in this case a tick) and the next generation is infected from early development Trypanosomes may be transferred occasionally from a female to suckling young [R P H]

Bibliography R P Hall *Protozoology* 1953 S H Hutner and A Lwoff (eds) *Biochemistry and Physiology of the Protozoa*, vol 2 1955, L H Hyman *The Invertebrates* vol 1 1940, R R Kudo, *Protozoology*, 4th ed 1954, A Lwoff and S H Hutner (eds), *Biochemistry and Physiology of the Protozoa*, vol 1, 1951

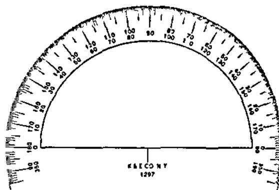
Protozoology

That branch of biology which deals with the Protozoa Medical protozoology is concerned primarily with parasites of man, veterinary protozoology with parasites of domestic animals [R P H]

Bibliography C F Craig and E C Faust, *Clinical Parasitology*, 5th ed, 1951, B B Morgan and P A Hawkins, *Veterinary Protozoology*, rev ed., 1952

Protractor

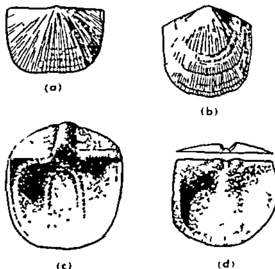
An instrument used to construct and measure angles formed by lines of a plane In its simplest form, it consists of half of a circular disk of metal or transparent material, with the bounding semicircle graduated in degrees (from 0° to 180°) The midpoint of the diameter of the semicircle is marked It serves as the vertex of angles constructed or measured A more complicated three-arm protractor is used in marine surveying *See* ANGLE, NAVIGATION [L M BL]



A protractor

Protremata

An order of specialized brachiopods of the class Articulata, with well developed articulation The dorsal and ventral valves are variously convex The exterior may be smooth striate costate or spiny



Billingsella coloradoensis (a) Dorsal exterior (Cambrian, Texas) (b) Ventral exterior (Cambrian Idaho) (c) Replica of dorsal interior (d) Replica of ventral interior (From W H Twenhofel and R R Shrock *Principles of Invertebrate Paleontology* McGraw Hill 1953)

Concentric growth lines may be well developed or obscure The pedicle is confined to the ventral valve and the delthyrium is more or less closed by the deltidium The cardinal area is well developed in this group Brachial supports are absent or rudi mentary The shell substance is either calcareous or composed of prismatic fibers This order de scended from the Palaeotremata through the super

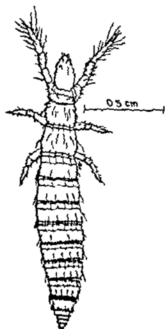
The appeared in the Cam

been re orded from

LATA (BRACHIOPODA) [K H]

Protura

An order of primitive wingless insects belonging to the subclass Apterygota These insects are less than 2 mm in length elongate fragile eyeless, and from pale amber to white in color They bear several characters which are unique in the class Insecta and which have led some authorities to place them in a separate class the Myrientomata Their meta morphosis is called anamorphosis that is a seg ment is added to the abdomen at each of the three molts a condition found in some of the lower Arthropoda Further they are the only insects in which antennae are absent The prothoracic legs functionally replace these structures These legs are long held anterolaterally to the head and bear l ng sensory hairs The mouthparts are composed of long slen ler stylets. One family the Eosentomi dae possesses spiracles and tracheae while the



Acerentulus barberi (From H E Ewing Ann Entomol Soc Am 33 31 497 1940)

other, the Acerentomidae lacks these structures Protura inhabit moist and decaying vegetation or moss and are most easily recovered through the use of a Berlese funnel Very little is known of their biology Originally described from Italy in 1907 they have since been found to be distributed throughout the world Although once considered rare they sometimes occur in large numbers in extractions from leaf mold *See* APTERYGOTA, INSECTA [H B MI]

Bibliography H E Ewing, The Protura of North America *Ann Entomol Soc Am* 33 495–551 1950, H B Mills, Catalogue of the Protura *Brooklyn Entomol Soc Bull* 27 125–130 1932

Proustite

A mineral having composition Ag_3AsS_3 and crystal lizing in the hexagonal system It occurs in pris matic crystals terminated by steep ditrigonal pyra mids but is more commonly massive or in disseminated grains There is good rhombohedral cleavage The hardness is 2–2 5 (Mohs scale) and the specific gravity is 5 55 The luster is adamantine and the color ruby red It is called light ruby silver in contrast to pyrargyrite dark ruby silver Proust ite is less common then pyrargyrite but the two minerals are found together in silver veins Noted localities are at Chañarcillo Chile Freiberg Ger many, Guanajuato Mexico, and Cobalt Ontario Canada *See* PYRARGYRITE, SILVER METALLURGY [C S HU]

Pseudoalleles

Closely linked genes with similar effects Pseudoal leles are exceptions to the general rule that genes which affect related functional processes in the or ganism are more or less randomly distributed throughout the chromosomes Pseudoalleles lie close together in the chromosome and are more or less closely related to one another in their functions As such they provide evidence that the spatial ar rangement of the genes in the chromosome plays a role in determining gene action. *See* GENE ACTION

Pseudoalleles are thought to be closely linked genes because crossing over which is the orderly process of recombination normally occurring be tween members of a pair of homologous chro mosomes occurs only rarely between them For example the probability of crossing over per gener ation between most known cases of pseudoallelic genes is seldom greater than 0 001 and more often it is 0 0001 or less *See* RECOMBINATION GENETIC

C s trans

het

pse

phe ding upon how the alleles are dis tributed between the pair of homologous chromo somes Thus if a and b symbolize the two reces sive pseudoallelic mutants then the heterozygote $a+/+b$ often has a mutant phenotype whereas the $ab/++$ heterozygote has a nonmutant or wild type phenotype The former is known as the *trans* het erozygote and the latter as the *cis* heterozygote

The difference in phenotype between the two is sometimes called the cis trans effect and is an important and perhaps the simplest example of the phenomenon of position effect Pseudoalleles which behave in this way are said to be position pseudoalleles One possible interpretation of this behavior is that such pseudoalleles control sequential steps in a chain of chemical reactions involving chemical compounds which for some reason are not readily diffusible from one chromosome to another On this basis the wild type alleles of *a* and *b* are expected to function more efficiently when together in the same chromosome as in $ab/++$ than when they are separated into opposite chromosomes as in $a+/+b$ (*see* ALLELE, MUTATION)

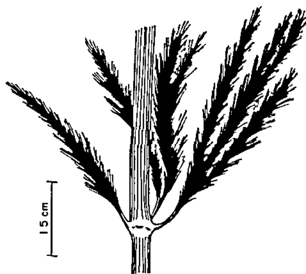
Cistron concept According to another terminology which is in use the genes *a* and *b* are said to belong to the same functional unit or cistron if the heterozygote $a+/+b$ is mutant in appearance If that heterozygote is wild type in phenotype *a* and *b* are said to belong to two different cistrons However a number of difficulties can arise, for example there can exist mutant alleles of a third pseudoallelic locus say *c*, such that $a+/+c$ and $b+/+c$ are mutant even though $a+/+b$ is wild type Thus the use of the cistron concept to define a genetic unit leads to ambiguities in some cases A crossing over analysis remains a more satisfactory way of clearly distinguishing between a multiple allelic and a pseudoallelic series The ultimate unit of crossing over has been termed the recon

Possible examples in man The pattern of inheritance of the Rhesus blood groups in man is sometimes considered to be an example of a pseudoallelic series composed of at least three loci which are designated *C D* and *E* it may also be considered to be a multiple allelic series of a single gene which is symbolized *R* for *Rhesus* The correct interpretation remains in doubt because of the difficulty involved in performing rigorous crossing over analyses in man *See* BLOOD GROUPS

Recombination between pseudoallelic genes has been shown to produce reciprocal recombinant types simultaneously as the result of a single exchange between the two loci In microorganisms and perhaps in higher organisms there is another process sometimes called gene conversion which resembles crossing over in its effect but which is *nonreciprocal and not necessarily associated* with recombination of neighboring genetic loci It is not clear whether gene conversion occurs between multiple alleles or between pseudoalleles When the relationship between gene conversion and crossing over is better understood it will be possible to define the limits of the ultimate genetic unit more precisely *See* GENE GENETICS [F R L.]

Pseudoborniales

An order of fossil plants found in Middle and Upper Devonian rocks The group is related to Sphenophyllales and includes a single family and two monotypic genera *Pseudobornia ursina* is known from Bear Island (*north of* Norway) and Germany



Pseudobornia node with three leaves (Modified from A G Nathorst)

Prosseria grandis is found in New York State Sphenopsid characters are more firmly established in this order than in Hyeniales

Pseudobornia the better known of the two has rhizomes and stems up to 6 cm wide and over a meter long The axes are jointed Larger axes bear two branches at a node and smaller ones have whorls of four leaves The leaves are 3 5–6 cm long short stalked palmately divided and they have laciniate margins Fertile shoots bear reduced leaves with sporangia on their lower sides *See* HYENIALES PALEOBOTANY, SPHENOPHYLLALES SPHENOPSIDA [H P B]

Pseudocoelomata

A group comprising the animal phyla Entoprocta and Aschelminthes in which there is an unlined space between the body wall and the digestive tract and other internal organs The space a relic of the embryonic blastocoele is not a true coelom or body cavity because it lacks a cellular lining or peritoneum In members of this group the digestive tract includes an anus protonephridia are either lacking or present and may or may not have flame cells *See* ACOELOMATA EUCOELOMATA [T I S]

Pseudomonadaceae

A family of bacteria of the suborder Pseudomonadineae The microorganisms are gram negative and nonsporeforming A few species in the genus *Pseudomonas* are pathogenic for plants animals and humans nearly all *Xanthomonas* species are pathogenic for plants and species in the genus *Acetomonas* are used in industrial microbiological processes for the production of vinegar gluconic acid and *l* sorbose Motile species invariably possess one or more flagella attached at the poles of the cells Nonmotile types resembling motile species are also included in the Pseudomonadaceae Photosynthetic pigments are not found but other types of pigments such as pyocyanin and fluorescein are common Many species require free access

to air for growth to occur, but others can also grow when air is excluded

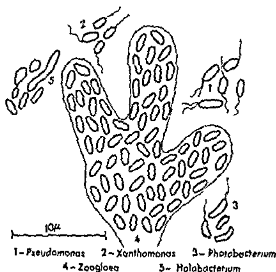
The majority of cells are rod shaped to ovoid and therefore distinguishable from the comma and spiral shaped cells of the Spirillaceae The cells, unlike those of the Caulobacteriaceae and the Siderocapsaceae are never attached to a substrate by a stalk or a mucilaginous excretion nor do they deposit iron or manganese in or on their cell walls, or capsules (thickened layer of slime surrounding each cell, or aggregation of cells) The inability of members of the Pseudomonadaceae to convert ammonia to nitrites, nitrites to nitrates, or to oxidize hydrogen, carbon monoxide methane or inorganic sulfur compounds differentiates them from the Nitrobacteriaceae, the Methanomonadaceae and the Thiobacteriaceae

Thirteen genera comprise the family, some of them little more than names to the majority of bacteriologists Five of these (*Aeromonas, Photobacterium, Kluyvera, Azotomonas,* and *Zymomonas*) are composed of bacteria that can decompose sugars in the absence of oxygen with the formation of characteristic end products The delineation of the other genera is at present based on characteristics of rather different sorts *Pseudomonas, Xanthomonas,* and *Acetomonas* in addition to the five genera previously mentioned are fairly well known to bacteriologists The remaining five genera (*Protaminobacter, Alginomonas, Mycoplana, Zoogloea,* and *Halobacterium*) are known only to specialists See BACTERIAL METABOLISM

Well known genera The eight well known genera are discussed in this section

Pseudomonas is the type genus after which the family is named It includes a variety of species—several cellulose decomposers, a few human and animal pathogens (disease agents), many plant pathogens and a number which are troublesome in various industries Most species are strictly aerobic which means that they must have free access to oxygen for growth to occur but several species can grow anaerobically, if nitrate or nitrite is present to serve as a hydrogen (or electron) acceptor Some representatives may produce gluconic 2 ketogluconic α-ketoglutaric pyruvic and succinic acids in glucose media other more or less distinctive products are polymers such as mannans and extracellular nucleic acids and water soluble pigments some of which have antibiotic properties The ability of some species to produce soluble pigments, such as the typical green fluorescence of cultures of certain *Pseudomonas* species makes them familiar to most bacter... bers of ... in soil ... tion.

Three species are widely known, although many others are recognized in books of determinative bacteriology One of these *Pseudomonas aeruginosa* is the type species It is also known as the blue pus organism because its presence in wounds leads to blue suppuration The color is imparted



Some genera of the Pseudomonadaceae (V B D Skerman)

by a soluble pigment, pyocyanin, which is produced by typical strains *Pseudomonas aeruginosa* is a pathogen for ma...

(see MELIOIDOSIS) *Pseudomonas fluorescens,* although a well described species in its own right, is sometimes a catchall for nonpathogenic isolates whose soluble pigments are not extractable with chloroform

Xanthomonas is primarily a genus of plant path...

...ture medium It is mainly by virtue of this latter characteristic that xanthomonads are differentiated from pseudomonads *Xanthomonas* cells are normally monotrichous, that is, possess a single polar flagellum They can oxidize a large variety of substances, but the products of their metabolism are not distinctive Except to plant pathologists, they are not well known

Acetomonas, comprising the polarly flagellated vinegar or acetic acid bacteria, is differentiated from *Pseudomonas* and *Xanthomonas* on the basis of the ability of its species to produce readily detectable amounts of acetic acid by the oxidation of ethanol (ethyl) alcohol) Nearly every species of *Pseudomonas* ... mean ... acetic ... that t... intolerance to acid of *Pseudomonas* species in contrast to the tolerance of acetobacters might be a more definitive distinction Although this suggestion has merit, a method for accomplishing the separation on this basis has not been proposed

The inclusion of *Acetomonas* instead of *Acetobacter* among the Pseudomonadaceae is a deviation from the classification in the seventh edition of *Bergey's Manual of Determinative Bacteriology*

The change is based on the work of Einar Leifson (1954) and J L Shimwell (1958) The old genus *Acetobacter* encompassed both polar and peritrichous species Inasmuch as the type species *A aceti* is peritrichous a new genus name was needed for the polarly flagellated species The name *Acetobacter* is retained for the peritrichously flagellated acetic acid bacteria It is included in the family Achromobacteraceae *See* ACHROMOBACTERACEAE

Acetobacters are used industrially for the production of vinegar and acetic acid Other common products of their oxidative activity are gluconic and 5 ketogluconic acids from glucose dihydroxy acetone from glycerol and sorbose from sorbitol

Aeromonas is composed of pseudomonads which physiologically resemble bacteria of the genus *Aerobacter*, family Enterobacteraceae of the Eubacteriales As presently constituted *Aeromonas* contains only four species three of which are known pathogens of fish and amphibians

Photobacterium differs from *Aeromonas* principally in that its species are luminescent their cultures emit light and thus glow in the dark Photobacteria are found in association with dead fish and other salt water animals

Azotomonas inhabits the soil Otherwise it is distinguished from *Aeromonas* because of its ability to fix atmospheric nitrogen

Kluyvera is the polar flagellate counterpart of the genus *Escherichia* family Enterobacteraceae order Eubacteriales *Kluyvera* species are especially notable for producing high yields of α keto glutaric acid Two possibly three species are presently recognized

Zymomonas differs from the other fermentative pseudomonads in that it causes a typical alcoholic fermentation of sugar The final alcohol concentration in the medium may reach 10% Hence *Zymomonas* is used in the production of alcoholic beverages such as pulque and beer

Little known genera The remaining five genera in this group are little known except to specialists

Protaminobacter is a genus which receives separate recognition because its members utilize alkyl amines as the sole source of carbon Otherwise it is not distinctive

Alginomonas is set apart because members of this genus are able to decompose alginic acid Most species are inhabitants of the seas

Mycoplana was probably selected as the name for this genus to signify a resemblance to fungi (myces) The resemblance is faint however residing in the propensity of some cells especially when young to show branching Phenol and similar aromatic compounds are utilized as the sole source of energy *Mycoplana* is probably widely distributed in soil

Zoogloea is the name of a genus of rod shaped bacteria which produces zoogloeal (gelatinous) masses in water containing decomposing organic matter Its members participate in the oxidation of sewage and industrial wastes *Zoogloea ramigera* is especially common in the flocs formed during sewage purification by the activated sludge process

Halobacterium is a genus of bacteria which requires the presence of at least 12% salt for growth Carotenoid pigments of orange to red shades are produced by some species Bacteria of this genus are found in tidal pools salt ponds salt seas and on salted fish and salted hides *See* BACTERIA TAXONOMY OF METHANOMONADACEAE, NITROBACTERIACEAE, PSEUDOMONAS AERUGINOSA, PSEUDOMONADINEAE, SCHIZOMYCETES, THIOBACTERIACEAE

[W C H]

Bibliography T Asai et al *Proc Imp Acad Japan* 32 488 1956 R S Breed et al *Bergey s Manual of Determinative Bacteriology* 7th ed., 1957 E Leifson The flagellation and taxonomy of species of *Acetobacter Antonie van Leeuwenhoek J Microbiol Serol* 20 102–110 1954 J L Shimwell Flagellation and taxonomy of *Acetobacter* and *Acetomonas Antonie van Leeuwenhoek J Microbiol Serol* 24 187–192 1958

Pseudomonadales

An order of bacteria of the class Schizomycetes Some of these organisms are parasitic and some pathogenic causing diseases of fish animals and man (cholera) The species are commonly found in the soil and in both fresh and salt waters Bacteria of the Pseudomonadales are often thought to be the most primitive among the Schizomycetes because of their lack of morphological differentiation and the ability possessed by many of them to exist on relatively simple primarily inorganic nutrients

Bacteria in this order have rigid cells which may be ovoid rod shaped comma shaped or spiral in form The cells usually 1 micron (μ) in diameter may occur singly in pairs and rarely in short or long chains A few species are exceptional in that diameters of 3–14 μ and lengths of 100 μ have been reported A majority of the species studied are gram negative (*see* GRAM S STAIN) Endo spores are not found in these bacteria Representative genera of this order are shown in Fig 1

Most of the members of the Pseudomonadales are motile by means of one or more whiplike appendages (flagella) which are attached to one or both ends or poles of individual cells In one genus (*Selenomonas*) the crescent shaped cells have tufts of flagella in their inner curvature (Fig 2) Motile forms in the orders Chlamydobacteriales and Hyphomicrobiales are also characterized by possession of polar flagella Chlamydobacteriales are distinguished from Pseudomonadales by the formation of trichomes permanent associations of dividing cells which may show differentiation into holdfast and reproductive cells The trichomes are often surrounded by sheaths composed of an organic matrix which may be impregnated with oxides of iron or manganese Differentiation between Pseudomonadales and Hyphomicrobiales depends upon the fact that multiplication of cells in the latter order is by budding or by budding and cell division whereas in the former only cell divi

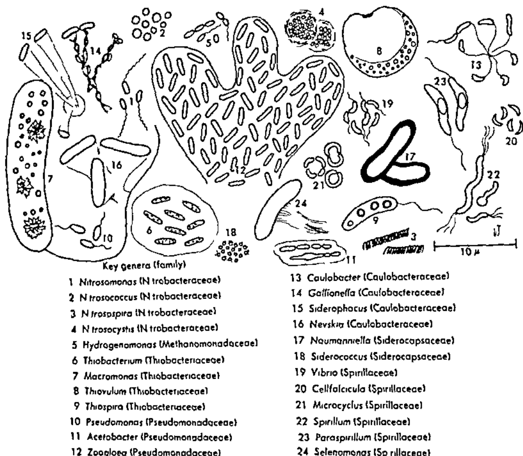


Fig 1 Representative genera of the Pseudomonadales (V B D Skerman)

sion (fission) is known to occur (*see* REPRODUCTION ANIMAL)

Many bacteria of this order form pigments Those which produce photosynthetic pigments are grouped together as a suborder called the Rhodobacteriineae (*see* PHOTOSYNTHESIS), the remainder, some of which make other types, are classified in the suborder Pseudomonadineae Pigments which are soluble in the culture medium are more common in certain genera of this suborder than elsewhere among the Schizomycetes The ability to produce pigments, especially of the sort soluble in the culture medium, is sometimes lost The classification and identification of such variants is often troublesome *See* BACTERIA, TAXONOMY OF, BACTERIAL ENDOSPORES, BACTERIAL MOTILITY, PSEUDOMONADINEAE, RHODOBACTERIINEAE, SCHIZOMYCETES

[W C H]

10 μ

Fig 2 Polar flagellate bacterium (*Pseudomonas saccharophila*) (Photomicrograph by Einar Leifson, Loyola University)

Pseudomonadineae

A suborder of bacteria of the order Pseudomonadales Some species are human, animal, or plant pathogens, others are organisms used in industrial fermentation processes The following families of the Pseudomonadales whose cells do not contain photosynthetic pigments are included in the suborder Nitrobacteraceae, Methanomonadaceae, Thiobacteriaceae, Pseudomonadaceae, Caulobacteraceae, Siderocapsaceae, and Spirillaceae They have the other characteristics of the order (*see* PSEUDOMONADALES) Many bacteria of this suborder produce nonphotosynthetic pigments, some

The change is based on the work of Einar Leifson (1954) and J L Shimwell (1958) The old genus *Acetobacter* encompassed both polar and peritrichous species Inasmuch as the type species *A aceti* is peritrichous a new genus name was needed for the polarly flagellated species The name *Acetobacter* is retained for the peritrichously flagellated acetic acid bacteria It is included in the family Achromobacteraceae *See* ACHROMOBACTERACEAE

Acetobacters are used industrially for the production *of vinegar and acetic acid Other common* products of their oxidative activity are gluconic and 5 ketogluconic acids from glucose dihydroxy acetone from glycerol and sorbose from sorbitol

Aeromonas is composed of pseudomonads which physiologically resemble bacteria of the genus *Aerobacter* family Enterobacteraceae of the Eubacteriales As presently constituted *Aeromonas* contains only four species three of which are known pathogens of fish and amphibians

Photobacterium differs from *Aeromonas* principally in that its species are luminescent their cultures emit light and thus glow in the dark Photobacteria are found in association with dead fish and other salt water animals

Azotomonas inhabits the soil Otherwise it is distinguished from *Aeromonas* because of its ability to fix atmospheric nitrogen

Kluyvera is the polar flagellate counterpart of the genus *Escherichia* family Enterobacteraceae order Eubacteriales *Kluyvera* species are especially notable for producing high yields of α keto glutaric acid Two possibly three species are presently recognized

Zymomonas differs from the other fermentative pseudomonads in that it causes a typical alcoholic fermentation of sugar The final alcohol concentration in the medium may reach 10% Hence *Zymomonas* is used in the production of alcoholic beverages such as pulque and beer

Little known genera The remaining five genera in this group are little known except to specialists

Protaminobacter is a genus which receives separate recognition because its members utilize alkyl amines as the sole source of carbon Otherwise it is not distinctive

Alginomonas is set apart because members of this genus are able to decompose alginic acid Most species are inhabitants of the seas

Mycoplana was probably selected as the name for this genus to signify a resemblance to fungi (myces) The resemblance is faint however residing in the propensity of some cells especially when young to show branching Phenol and similar aromatic compounds are utilized as the sole source *of energy Mycoplana* is probably widely distributed in soil

Zoogloea is the name of a genus of rod shaped bacteria which produces zoogloeal (gelatinous) masses in water containing decomposing organic matter Its members participate in the oxidation of sewage and industrial wastes *Zoogloea ramigera* is especially common in the flocs formed during sewage purification by the activated sludge process

Halobacterium is a genus of bacteria which requires the presence of at least 12% salt for growth Carotenoid pigments of orange to red shades are produced by some species Bacteria of this genus are found in tidal pools salt ponds salt seas and on salted fish and salted hides *See* BACTERIA TAXONOMY OF METHANOMONADACEAE NITROBACTERIACEAE PSEUDOMONAS AERUGINOSA PSEUDOMONADINEAE, SCHIZOMYCETES THIOBACTERIACEAE

[W C H]

Bibliography T Asai et al *Proc Imp Acad Japan* 32 488 1956 R S Breed et al *Bergey's Manual of Determinative Bacteriology* 7th ed 1957 E Leifson The flagellation and taxonomy of species of *Acetobacter Antonie van Leeuwenhoek J Microbiol Serol* 20 102–110 1954 J L Shimwell Flagellation and taxonomy of *Acetobacter* and *Acetomonas Antonie van Leeuwenhoek J Microbiol Serol* 24 187–192 1958

Pseudomonadales

An order of bacteria of the class Schizomycetes Some of these organisms are parasitic and some pathogenic causing diseases of fish animals and man (cholera) The species are commonly found in the soil and in both fresh and salt waters Bacteria *of the Pseudomonadales are often thought to be the* most primitive among the Schizomycetes because of their lack of morphological differentiation and the ability possessed by many of them to exist on relatively simple primarily inorganic nutrients

Bacteria in this order have rigid cells which may be ovoid rod shaped comma shaped or spiral in form The cells usually 1 micron (μ) in diameter may occur singly in pairs and rarely in short or long chains A few species are exceptional in that diameters of 3–14 μ and lengths of 100 μ have been reported A majority of the species studied are gram negative (*see* GRAM S STAIN) Endospores are not found in these bacteria Representative genera of this order are shown in Fig 1

Most of the members of the Pseudomonadales are motile by means of one or more whiplike appendages (flagella) which are attached to one or both ends or poles of individual cells In one genus (*Selenomonas*) the crescent shaped cells have tufts of flagella in their inner curvature (Fig 2) Motile forms in the orders Chlamydobacteriales and Hyphomicrobiales are also characterized by possession of polar flagella Chlamydobacteriales are distinguished from Pseudomonadales by the formation of trichomes permanent associati ns of dividing cells which may show differentiation into holdfast and reproductive cells The trichomes often surrounded by sheaths composed of an ganic matrix which may be impregnated oxides of iron or manganese Differentiatic tween Pseudomonadales and Hyphomicrob pends upon the fact that multipl cati n the latter order is by budding or by l cell division whereas in the f rmer r l

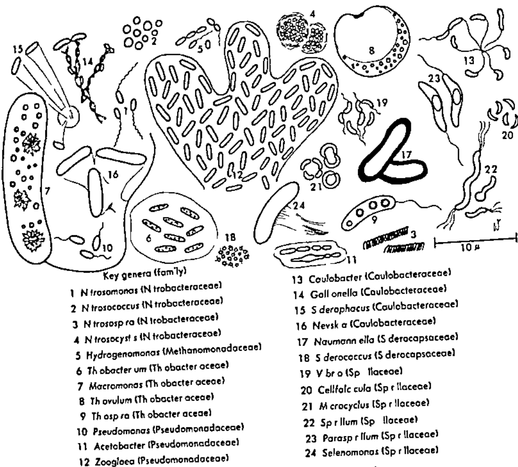


Fig 1 Representative genera of the Pseudomonadales (V B D Skerman)

sion (fission) is known to occur (see REPRODUCTION ANIMAL)

Many bacteria of this order form pigments Those which produce photosynthetic pigments are grouped together as a suborder called the Rhodobacteriineae (see PHOTOSYNTHESIS), the remainder some of which make other types are clas [illegible]

elsewhere among the Schizomycetes The ability to produce pigments especially of the sort soluble in the culture medium is sometimes lost The classification and identification of such variants is often troublesome See BACTERIA TAXONOMY OF, BACTERIAL ENDOSPORES, BACTERIAL MOTILITY PSEUDOMONADINEAE, RHODOBACTERIINEAE, SCHIZOMYCETES [W C H]

Fig 2 Polar flagellate bacterium (Pseudomonas saccharophila (Photomicrograph by Einar Leifson Loyola University)

Pseudomonadineae

A suborder of bacteria of the order Pseudomonadales Some species are human animal or plant pathogens others are organisms used in industrial fermentation processes The following families of the Pseudomonadales whose cells do not contain photosynthetic pigments are included in the suborder Nitrobacteraceae Methanomonadaceae Thiobacteriaceae Pseudomonadaceae Caulobacteraceae Siderocapsaceae and Spirillaceae They have the other characteristics of the order (see PSEUDOMONADALES) Many bacteria of this suborder produce nonphotosynthetic pigments some

The change is based on the work of Einar Leifson (1954) and J L. Shimwell (1958) The old genus *Acetobacter* encompassed both polar and peritrichous species Inasmuch as the type species *A aceti* is peritrichous a new genus name was needed for the polarly flagellated species The name *Acetobacter* is retained for the peritrichously flagellated acetic acid bacteria It is included in the family Achromobacteraceae *See* ACHROMOBACTERACEAE

Acetobacters are used industrially for the production of vinegar and acetic acid Other common products of their oxidative activity are gluconic and 5 ketogluconic acids from glucose dihydroxy acetone from glycerol and sorbose from sorbitol

Aeromonas is composed of pseudomonads which physiologically resemble bacteria of the genus *Aerobacter*, family Enterobacteraceae of the Eubacteriales As presently constituted *Aeromonas* contains only four species three of which are known pathogens of fish and amphibians

Photobacterium differs from *Aeromonas* principally in that its species are luminescent their cultures emit light and thus glow in the dark Photobacteria are found in association with dead fish and other salt water animals

Azotomonas inhabits the soil Otherwise it is distinguished from *Aeromonas* because of its ability to fix atmospheric nitrogen

Kluyvera is the polar flagellate counterpart of the genus *Escherichia* family Enterobacteraceae order Eubacteriales *Kluyvera* species are especially notable for producing high yields of α keto glutaric acid Two possibly three species are presently recognized

Zymomonas differs from the other fermentative pseudomonads in that it causes a typical alcoholic fermentation of sugar The final alcohol concentration in the medium may reach 10% Hence *Zymomonas* is used in the production of alcoholic beverages such as pulque and beer

Little known genera The remaining five genera in this group are little known except to specialists

Protaminobacter is a genus which receives separate recognition because its members utilize alkyl amines as the sole source of carbon Otherwise it is not distinctive

Alginomonas is set apart because members of this genus are able to decompose alginic acid Most species are inhabitants of the seas

Mycoplana was probably selected as the name for this genus to signify a resemblance to fungi (myces) The resemblance is faint however residing in the propensity of some cells especially when young to show branching Phenol and similar aromatic compounds are utilized as the sole source of energy *Mycoplana* is probably widely distributed in soil

Zoogloea is the name of a genus of rod shaped bacteria which produces zoogloeal (gelatinous) masses in water containing decomposing organic matter Its members participate in the oxidation of sewage and industrial wastes *Zoogloea ramigera* is especially common in the flocs formed during sewage purification by the activated sludge process

Halobacterium is a genus of bacteria which requires the presence of at least 12% salt for growth Carotenoid pigments of orange to red shades are produced by some species Bacteria of this genus are found in tidal pools salt ponds salt seas and on salted fish and salted hides *See* BACTERIA TAXONOMY OF, METHANOMONADACEAE, NITROBACTERIACEAE, PSEUDOMONAS AERUGINOSA, PSEUDOMONADINEAE SCHIZOMYCETES, THIOBACTERIACEAE

[W C H]

Bibliography T Asai et al *Proc Imp Acad Japan* 32 488 1956, R S Breed et al *Bergey's Manual of Determinative Bacteriology*, 7th ed 1957, E Leifson The flagellation and taxonomy of species of *Acetobacter*, *Antonie van Leeuwenhoek J Microbiol Serol*, 20 102–110 1954, J L Shimwell Flagellation and taxonomy of *Acetobacter* and *Acetomonas*, *Antonie van Leeuwenhoek*, *J Microbiol Serol* 24 187–192 1958

Pseudomonadales

An order of bacteria of the class Schizomycetes Some of these organisms are parasitic and some pathogenic causing diseases of fish animals and man (cholera) The species are commonly found in the soil and in both fresh and salt waters Bacteria of the Pseudomonadales are often thought to be the most primitive among the Schizomycetes because of their lack of morphological differentiation and the ability possessed by many of them to exist on relatively simple primarily inorganic nutrients

Bacteria in this order have rigid cells which may be ovoid rod shaped comma shaped or spiral in form The cells usually 1 micron (μ) in diameter may occur singly in pairs and rarely in short or long chains A few species are exceptional in that diameters of 3–14 μ and lengths of 100 μ have been reported A majority of the species studied are gram negative (*see* GRAM'S STAIN) Endospores are not found in these bacteria Representative genera of this order are shown in Fig 1

Most of the members of the Pseudomonadales are motile by means of one or more whiplike appendages (flagella) which are attached to one or both ends or poles of individual cells In one genus (*Selenomonas*) the crescent shaped cells have tufts of flagella in their inner curvature (Fig 2) Motile forms in the orders Chlamydobacteriales and Hyphomicrobiales are also characterized by possession of polar flagella Chlamydobacteriales are distinguished from Pseudomonadales by the formation of trichomes permanent associations of dividing cells which may show differentiation into holdfast and reproductive cells The trichomes are often surrounded by sheaths composed of an organic matrix which may be impregnated with oxides of iron or manganese Differentiation between Pseudomonadales and Hyphomicrobiales depends upon the fact that multiplication of cells in the latter order is by budding or by budding and cell division whereas in the former only cell divi

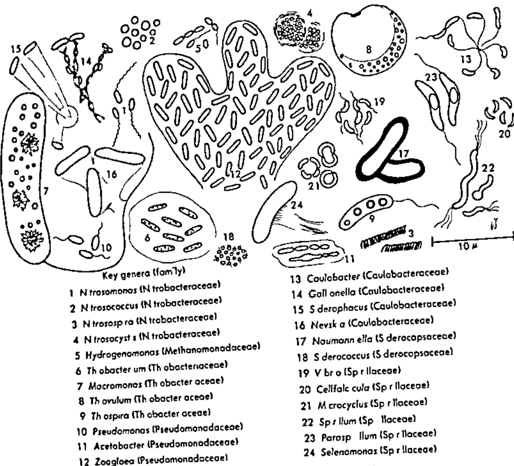


Fig 1 Representative genera of the Pseudomonadales (V B D Skerman)

sion (fission) is known to occur (*see* REPRODUCTION ANIMAL)

Many bacteria of this order form pigments Those which produce photosynthetic pigments are grouped together as a suborder called the Rhodobacteriineae (*see* PHOTOSYNTHESIS), the remainder some of which make other types are classified in the suborder Pseudomonadineae Pigments

the culture medium is sometimes lost The classification and identification of such variants is often troublesome *See* BACTERIA TAXONOMY OF, BACTERIAL ENDOSPORES BACTERIAL MOTILITY, PSEUDOMONADINEAE RHODOBACTERIINEAE SCHIZOMYCETES [W C H]

10 μ

Fig 2 Polar flagellate bacterium (*Pseudomonas saccharophila* (Photomicrograph by Einar Leifson Loyola University)

Pseudomonadineae

A suborder of bacteria of the order Pseudomonadales Some species are human animal or plant pathogens others are organisms used in industrial fermentation processes The following families of the Pseudomonadales whose cells do not contain photosynthetic pigments are included in the suborder Nitrobacteraceae Methanomonadaceae Thiobacteriaceae Pseudomonadaceae Caulobacteraceae Siderocapsaceae and Spirillaceae They have the other characteristics of the order (*see* PSEUDOMONADALES) Many bacteria of this suborder produce nonphotosynthetic pigments some

of which are soluble in the culture medium *See* PSEUDOMONADACEAE

The first four families listed above are composed of rods which are normally straight or coccoid and which display few or no distinctive morphological features Their differentiation therefore is based on unique physiological characteristics The remaining three families have cells which are curved spirally twisted coccoid ellipsoidal or rodlike Their differentiation is primarily morphological The principal characteristic of the Spirillaceae is the presence of curved or spirally twisted cells which do not have stalks and are not embedded in a capsule a thickened layer of slime which surrounds each cell or aggregation of cells The best known members of this family are the human and animal pathogens *Vibrio comma* (causative agent of Asiatic cholera) and *Vibrio fetus* (causative agent of abortion in cattle and sheep) Unlike the free living cells of the Spirillaceae those of the Caulobacteraceae normally are attached to a substrate by branching or unbranching stalks Floating forms are also known Bacteria of the family Siderocapsaceae do not possess stalks but their cells are usually embedded in a thick mucilaginous capsule in which iron or manganese compounds may be deposited They may be free-living or attached to surfaces of submerged objects Neither the caulobacteria nor members of the Siderocapsaceae are familiar to most bacteriologists *See* CAULOBACTERACEAE SIDEROCAPSACEAE, SPIRILLACEAE

Among the free-living straight rod forms are some which are able to oxidize ammonia to nitrites or nitrites to nitrates these belong to the Nitrobacteraceae Others which oxidize methane hydrogen or carbon monoxide are assigned to the Methanomonadaceae The Thiobacteriaceae embrace those forms which oxidize sulfur compounds frequently with a deposit of free sulfur globules or crystals within the cells or in the surrounding medium The bacteria of these three families also are little known to the majority of bacteriologists *See* METHANOMONADACEAE NITROBACTERACEAE THIOBACTERIACEAE

Rod forms which belong to the Pseudomonadineae but are not assignable to any one of the families described above are relegated to the Pseudomonadaceae This family is physiologically heterogeneous because it encompasses both oxidative and fermentative bacteria The oxidative types are best known primarily because of their economic importance Among them are a majority of the plant pathogenic bacteria and the human and animal pathogens *Pseudomonas pseudomallei* and *Pseudomonas aeruginosa* Also included are several species capable of carrying out industrial fermentations (vinegar 2 ketogluconic acid and sorbose) and others which cause damage to dairy and poultry products A salient characteristic of many species especially in the genus *Pseudomonas* is the ability to produce pigments which are soluble in the culture medium *See* INDUSTRIAL MICROBIOLOGY [W C H]

Bibliography R S Breed et al., *Bergey's Manual of Determinative Bacteriology* 7th ed., 1957

Pseudomonas aeruginosa

A species of the bacterial genus *Pseudomonas* which is also known as *Pseudomonas pyocyanae* or *Bacterium pyocyaneum* It is a nonsporeforming motile gram negative rod with 1–3 flagella It produces a greenish yellow pigment called fluorescein and a bluish green antibacterial pigment known as pyocyanin *Pseudomonas aeruginosa* is found in the intestinal tract and sometimes on the skin of man and animals as well as in sewage and polluted water It may cause urinary and wound infections and occasionally infant diarrhea meningitis or septicemia Infected wounds produce a peculiar blue pus because of the pigments of the bacterium *See* BACTERIOLOGY MEDICAL PSEUDOMONADACEAE [A J W]

Pseudophyllidea

An order of tapeworms of the subclass Cestoda parasitic in the intestine of all classes of vertebrates Typically the head is simple in structure with two groovelike attachment organs (Fig 1*a*) the bothria

Most pseudophyllideans are segmented and polyzoic with replication of the reproductive systems although there are a number which do not show such replication and are monozoic The genital openings are typically in the midline rather than lateral and there is usually a uterine pore in the midline from which embryos are discharged (Fig 1*b*) *Dibothriocephalus latus* the broad or fish tapeworm of man and certain piscivorous mammals is a pseudophyllidean In man this worm sometimes precipitates a pernicious anemia by competing with the host for vitamin B_{12}

After leaving the intestine of the mammalian host the ciliated embryo or coracidium of *D latus* escapes from its shell and must be eaten by an arthropod a copepod of the genus *Cyclops* or *Diaptomus* In the copepod the embryo develops into

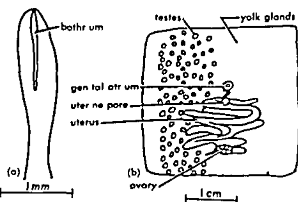


Fig 1 (a) Scolex of D bothriocephalus (b) A pseudophyllidean segment diagrammatic

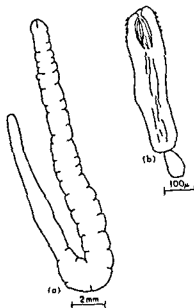


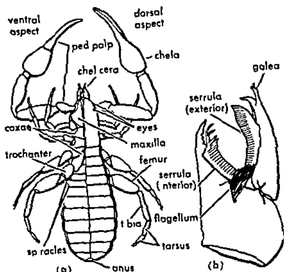
Fig 2 D bothriocephalus (a) Procerco d larva (b) Ple rocerco d larva

a procercoid (Fig 2a) For development to con tinue the copepod must be eaten by a fish in which the worm develops into a plerocercoid larva (Fig 2b) If this plerocercoid is eaten by the mam malian host, the worm rapidly grows to sexual ma turity sometimes attaining a length of more than 30 ft Larval pseudophyllideans are occasionally found as parasites in the extraintestinal tissues of man producing a condition known as spargano sis *See* CESTODA SPARGANOSIS [C P R]

Bibliography L. A Hyman *The Invertebrates*, vol 2, 1951 R A Wardle and J A McLeod *The Zoology of Tapeworms* 1952

Pseudoscorpionida

An order of terrestrial Arachnida having the general appearance of miniature scorpions without the postabdomen and sting These animals are also known as the Chelonethida The body length is seldom greater than 5 0 mm Typically each finger of the anterior appendages or chelicerae has a serrula composed of a row of ligulate plates Ducts of silk glands open near the end of the movable finger often in connection with a simple or branched spinneret or galea The second pair of appendages or palpi are large and conspicuous usually with glands that discharge venom through a terminal tooth on one or both of the chelal fingers The four pairs of legs are ambulatory Eggs and larval young are frequently carried on the underside of the abdomen of the female Pseudoscorpions feed chiefly on small arthropods and although frequently found on birds mammals and insects are considered nonparasitic Pseudoscorpions are common in the nests of mammals birds and social insects in woody debris and forest litter under stones and in crevices in the bark of trees A few species are found in barns and chicken houses and the virtually cosmopolitan species *Chelifer cancroides*, is frequently found in libraries and dwellings Tertiary fossil pseudoscorpions



(a) Diagram of a pseudoscorpion (b) Chelicera of male pseudoscorpion (From H S Pratt, *A Manual of the Common Invertebrate Animals*, rev ed, McGraw Hill 1951)

from Baltic amber are very similar to modern forms [C C H O]

Bibliography M Beier, *Pseudoscorpionidea* in F E Schulze and W Kukenthal (eds), *Das Tierreich*, vols 57–58, 1932, J C Chamberlain The arachnid order Chelonethida, *Stanford Univ Publs*, *Biol Sci*, 7(1) 1–284 1931, C C. Hoff List of the pseudoscorpions of North America north of Mexico *Am Museum Novitates*, 1875 1–50, 1958

Pseudotuberculosis

A disease of rodents and birds caused by a bacterium *Pasteurella pseudotuberculosis* The disease is occasionally transmitted to man

Pasteurella pseudotuberculosis is a large, pleomorphic flagellated organism, occurring in chains It is motile at 18–22°C and gram negative since it stains red with Gram's stain Occasionally the ends of the organism will stain more intensely than the center (bipolar staining)

The organism will grow on a medium containing bile salts and in an amino acid solution without accessory growth factors The organism hydrolyzes urea and does not produce gas in a carbohydrate medium

TL

roc · · ·

hy

infection has been reported as an acute fatal septicemia in 15 cases The organisms may localize in mesenteric lymph nodes of the ileocecal region and cause acute appendicitis and gastrointestinal symptoms Tetracycline drugs such as chlortetra

cycline tetracycline and oxytetracycline are probably effective

Diagnosis is made by bacteriological examination since the disease cannot be distinguished clinically or anatomically from typhoid paratyphoid tularemia or tuberculosis Bacteriological differentiation must be made between *Pasteurella pseudotuberculosis* and *Pasteurella pestis* For taxonomy *see* BRUCELLACEAE for motility *see* BACTERIAL MOTILITY *see also* SEPTICEMIA [K F ME]

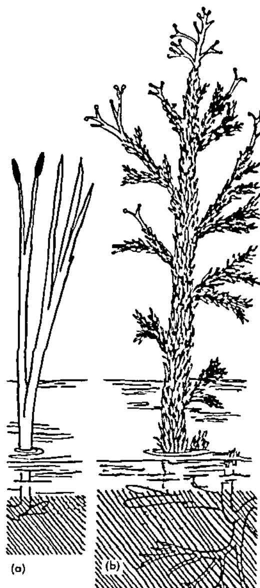
Psilomelane

A basic oxide of barium and manganese with the idealized chemical composition $BaMnMn_8O_{16}(OH)_4$ X ray structural studies have shown that psilomelane is orthorhombic in crystallization The mineral itself is not well crystallized and typically occurs as fine grained masses and crusts with a botryoidal or reniform structure The color is iron black to dark steel gray The hardness is about 5½ on Mohs scale and the specific gravity is 4 71 Psilomelane often occurs admixed with other manganese oxides chiefly pyrolusite and with clay and hydrated iron oxides The recognition of psilomelane and the proper use of the name has been attended by much confusion The name formerly was used in part in a generic sense for ill defined hard fine-grained manganese oxides that often contained little or no barium and the true status of many psilomelanelike minerals described in the literature or preserved in museum collections still remains uncertain Psilomelane is a secondary mineral formed under surface or near surface conditions of temperature and pressure [C FR]

Psilophytales

An order of fossil plants (subphylum Psilopsida) found in rocks of Silurian and Devonian age The plant body shows a low degree of organ differentiation and consists mainly of a rhizome bearing simply branched rhizoids and aerial stems The stems are naked or spiny with stomata on the surface or with small spiny leaves Sporangia (spore producing structures) are borne at the tips of branches The psilophytes apparently grew near water in swamps and marshes as do the rushes of today which they resemble in appearance and size It is assumed that they were periodically covered by flood water causing large masses to become embedded in the sand Silica dissolved in the water slowly penetrated the tissues so that even the cell walls have been remarkably preserved

Morphology Some Psilophytales resemble a *fern plant* others are similar to horsetails The stele is a protostele a stele with a solid central core or xylem (water conducting tissue) completely surrounded by a band of phloem (food conducting tissue) and having no pith Although sporophytes (spore-producing generation) have been discovered no gametophytes (gamete-producing generation) have been found Nevertheless it is thought that these plants must have had an alternation of generations



Sketch of psilophyte plants growing in shallow pool with runners in earth (a) *Rhynia* a leafless plant with spore cases at tips of fertile shoots (b) *Asteroxylon* covered with short scaly leaves except at tips of some branches which bear terminal sporangia (*Adapted from R C Moore Introduction to Historical Geology 2d ed McGraw-Hill 1958*)

Evolutionary significance The discovery of the Psilophytales has strengthened the theory that the vascular plants (plants with specialized water and food conducting tissues) evolved directly from the algae rather than from the Bryophyta The latter do not appear in the fossil record until the Carboniferous period many millions of years later *See* EMBRYOPHYTA PALEOBOTANY PLANT KINGDOM PSILOPSIDA [P D S]

Psilophytineae

The sole class of the plant subphylum Psilopsida The class is subdivided into two orders Psilophytales and Psilotales See PSILOPHYTALES PSILOTALES *see also* PLANT KINGDOM PSILOPSIDA [P D S]

Psilopsida

One of the four subphyla of the plant phylum Tracheophyta The members of this subphylum are the most primitive vascular plants known They have a dichotomously (forked) branching stem but no true roots or leaves The plant body is structurally simple The stele is mainly of the type known as a protostele a stele with a solid central core of xylem (water-conducting tissue) completely surrounded by a band of phloem (food conducting tissue) and having no pith The wood is extremely simple and homogeneous None of the Psilopsida is believed to possess a vascular cambium The entire body con

only one type (homosporous) The subphylum consists of the single class Psilophytineae (see PSILOPHYTINEAE) This class is subdivided into two orders Psilophytales and Psilotales See PSILOPHYTALES PSILOTALES *see also* PLANT KINGDOM [P D S]

Bibliography See EMBRYOPHYTA

Psilotales

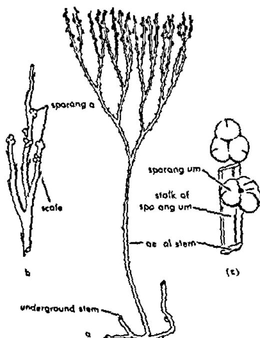
An order of the plant subphylum Psilopsida consisting of two living genera *Psilotum* with two species and *Tmesipteris* with a single species They have tiny emergences on the stems which may be regarded as primitive leaves The sporangia (spore producing structures) are borne laterally in the axils of these leaves instead of being terminal as in the fossil Psilophytales The underground part (rhizome) has a protostele (a solid core of vascular tissue without a pith) but the branches have a central pith hence the stele of the aerial portion is a siphonostele These plants have an alternation of generations Both sporophytes (spore-producing generation) and gametophytes (sex cell producing generation) have been studied the latter are subterranean dichotomously (forked) branching bodies resembling pieces of rhizomes.

Psilotum is tropical and subtropical in distribution being found in Florida Bermuda and Hawaii In temperate regions it is often grown in greenhouses In this genus the sporophyte is made up of a dichotomous green aerial shoot arising from an underground stem The plant does not have roots and the aerial branches bear paired scales rather than leaves Antheridia (male sex organs) and archegonia (female sex organs) similar to those of the Bryophyta cover the gametophytes

Tmesipteris occurs in Australia New Zealand New Caledonia and as far north as the Philippine Islands *Tmesipteris* is commonly an epiphyte (perched upon another plant) and is similar to *Psilotum* in that it is also rootless and the branches if any are dichotomous Two lobed sporangia occur in the axils of the leaves

The Psilotales are regarded as the living descendants of the ancient group Psilophytales now known only as fossils *See* PSILOPHYTALES PSILOPSIDA *see also* PLANT KINGDOM [P D S]

Bibliography See EMBRYOPHYTA



Psilotum a primitive living vascular plant (a) Habit of plant with rootless underground stem and aerial dichotomously branched stem with numerous scale leaves and three lobed sporangia (b) Enlarged portion of branch tips showing scale leaves and sporangia (c) Three lobed sporangia attached to the stalk the upper figure shows the line of dehiscence in each lobe

Psittaciformes

An order of birds containing the single family Psittacidae the parrots This group with over 300 species is predominantly tropical but has extended as far north as the central United States (the extinct Carolina parakeet *Conuropsis carolinensis*) and as far south as Tierra del Fuego (Chilean paroquet *Microsittace ferruginea*) Only one species the gray breasted parakeet (*Myopsitta monachus*) of South America is known to build its own nest of sticks All other parrots nest in holes or cavities Parrots are highly popular as cagebirds many have brilliantly colored plumage and some have exceptional powers of vocal mimicry Frequently kept are the New World macaws (*Ara*) and amazons (*Amazona*) the African love birds (*Agap*

pap

dul

ferentiated one of uncertain relationship *See* AVES PSITTACOSIS [K C P]

Psittacosis

An infection of man derived from birds of 96 species of which 57 are psittacine as well as an infection in psittacine birds both caused by the same

cycline tetracycline and oxytetracycline are probably effective

Diagnosis is made by bacteriological examination, since the disease cannot be distinguished clinically or anatomically from typhoid paratyphoid tularemia, or tuberculosis. Bacteriological differentiation must be made between *Pasteurella pseudotuberculosis* and *Pasteurella pestis*. For taxonomy see BRUCELLACEAE for motility see BACTERIAL MOTILITY see *also* SEPTICEMIA [K.F.ME.]

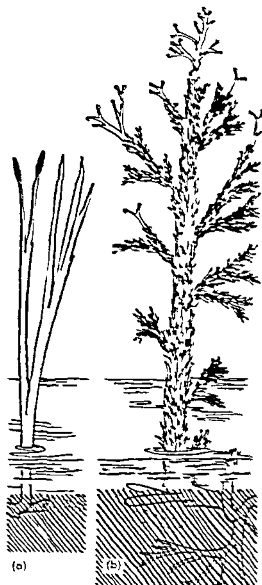
Psilomelane

A basic oxide of barium and manganese with the idealized chemical composition $BaMnMn_8O_{16}(OH)_4$. X-ray structural studies have shown that psilomelane is orthorhombic in crystallization. The mineral itself is not well crystallized and typically occurs as fine-grained masses and crusts with a botryoidal or reniform structure. The color is iron black to dark steel gray. The hardness is about $5\frac{1}{2}$ on Mohs scale and the specific gravity is 4.71. Psilomelane often occurs admixed with other manganese oxides, chiefly pyrolusite and with clay and hydrated iron oxides. The recognition of psilomelane and the proper use of the name has been attended by much confusion. The name formerly was used in part in a generic sense for ill-defined hard fine-grained manganese oxides, that often contained little or no barium, and the true status of many psilomelanelike mineral described in the literature or preserved in museum collections still remains uncertain. Psilomelane is a secondary mineral formed under surface or near surface conditions of temperature and pressure [C.FR.]

Psilophytales

An order of fossil plants (subphylum Psilopsida) found in rocks of Silurian and Devonian age. The plant body shows a low degree of organ differentiation and consists mainly of a rhizome bearing simply branched rhizoid and aerial stems. The stems are naked or spiny with stomata on the surface or with small spiny leaves. Sporangia (spore-producing structures) are borne at the tips of branches. The psilophytes apparently grew near water in swamp and marshes, as do the rushes of today which they resemble in appearance and size. It is assumed that they were periodically covered by flood water causing large masses to become embedded in the sand. Silica dissolved in the water slowly penetrated the tissues so that even the cell wall have been remarkably preserved.

Morphology Some Psilophytales resemble a fern plant others are similar to horsetails. The stele is a protostele a stele with a solid central core or xylem (water-conducting tissue) completely surrounded by a band of phloem (food-conducting tissue) and having no pith. Although sporophytes (spore-producing generation) have been discovered no gametophytes (gamete-producing generation) have been found. Nevertheless, it is thought that these plants must have had an alternation of generations.



Sketch of psilophyte plants growing in shallow pool with runners in earth. (a) *Rhynia* a leafless plant with spore cases at tips of fertile shoots. (b) *Asteroxylon* covered with short scaly leaves except at tips of some branches which bear terminal sporangia. (*Adapted from R. C. Moore Introduction to Historical Geology 2d ed., McGraw-Hill 1955*)

Evolutionary significance. The discovery of the Psilophytales has strengthened the theory that the vascular plants (plants with specialized water and food-conducting tissues) evolved directly from the algae rather than from the Bryophyta. The latter do not appear in the fossil record until the Carboniferous period many millions of years later. See EMBRYOPHYTA; PALEOBOTANY; PLANT KINGDOM; PSILOPSIDA. [P.D.S.]

Psilophytineae

The sole class of the plant subphylum Psilopsida. The class is subdivided into two orders, Psilophytales and Psilotales. See PSILOPHYTALES; PSILOTALES; see *also* PLANT KINGDOM; PSILOPSIDA. [P.D.S.]

Psilopsida

One of the four subphyla of the plant phylum Tracheophyta The members of this subphylum are the most primitive vascular plants known They have a dichotomously (forked) branching stem but no true roots or leaves The plant body is structurally simple The stele is mainly of the type known as a protostele a stele with a solid central core of xylem (water conducting tissue) completely surrounded by a band of phloem (food conducting tissue) and having no pith The wood is extremely simple and *homogeneous* None of the P ilopsida is believed to pos ess a vascular cambium The entire body con

only one type (homosporous) The subphylum consists of the single class Psilophytineae (see PSILOPHYTINEAE) This class is subdivided into two orders Ps lophytales and Psilotales *See* PSILOPHYTALES PSILOTALES *see also* PLANT KINGDOM [P D S]

Bibliography *See* EMBRYOPHYTA

Psilotales

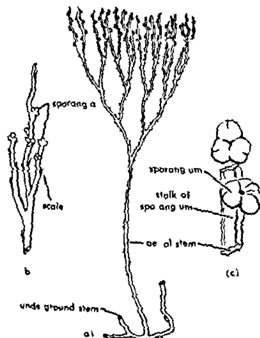
An order of the plant subphylum Psilopsida consisting of two living genera *Psilotum* with two species and *Tmesipteris* with a single species They have tiny emergences on the stems which may be *regarded* as primitive leaves The sporangia (spore producing structures) are borne laterally in the axils of these leaves instead of being terminal as in the fossil Psilophytales The underground part (rhizome) has a protostele (a solid core of vascular tissue without a pith) but the branches have a central pith hence the stele of the aerial portion is a siphonostele These plants have an alternation of generations Both sporophytes (spore producing generation) and gametophytes (sex cell producing generation) have been studied the latter are subterranean dichotomously (forked) branching bodies resembling pieces of rhizomes

Psilotum is tropical and subtropical in distribution being found in Florida Bermuda and Hawaii In temperate regions it is often grown in green houses In this genus the sporophyte is made up of a dichotomous green aerial shoot arising from an underground stem The plant does not have roots and the aerial branches bear paired scales *rather than leaves* Antheridia (male sex organs) and archegonia (female sex organs) similar to those of the Bryophyta cover the gametophytes

Tmesipteris occurs in Australia New Zealand New Caledonia and as far north as the Philippine Islands *Tmesipteris* is commonly an epiphyte (perched upon another plant) and is similar to *Psilotum* in that it is also rootless and the branches if any are dichotomous Two-lobed sporangia occur in the axils of the leaves

The Psilotales are regarded as the living descendants of the ancient group Psilophytales now known only as fossils *See* PSILOPHYTALES *PSILOPSIDA* *see also* PLANT KINGDOM [P D S]

Bibliography See EMBRYOPHYTA



Psilotum a primitive living vascular plant (a) Habit of plant with rootless underground stem and aerial dichotomously branched stem w h numerous scale leaves and three lobed spo ang a (b) Enlarged portion of branch tips showing scale leaves and sporangia (c) Three lobed *sporangia* attached to the stalk the upper figure shows the line of dehiscence in each lobe

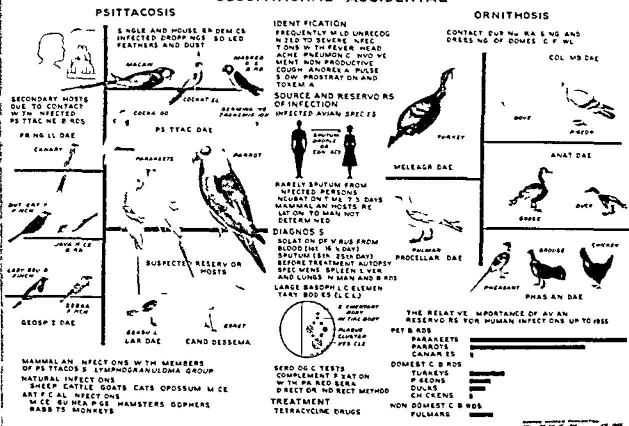
Psittaciformes

An order of birds containing the single family Psittacidae the parrots This group with over 300 species is predominantly tropical but has extended as far north as the central United States (the extinct Carolina *parakeet* *Conuropsis carolinensis*) and as far south as Tierra del Fuego (Chilean paroquet *Microsittace ferruginea*) Only one species the —

t ies or cavities Parrots are highly popular as cagebirds many have brilliantly colored plumage and some have exceptional powers of *vocal* mimicry Frequently kept are the New World macaws (*Ara*) and amazons (*Amazona*) the African love birds (*Agapornis*) and gray parrot (*Psittacus*) and most popular of all the budgerigar (*Melopsittacus undulatus*) of Australia The order is a strongly differentiated *one* of uncertain relationship *See* AVES PSITTACOSIS [K C P]

Psittacosis

In infect
cies of w
tion in p
e same



Epidemiology of psittacosis-ornithosis (George Hooper Foundation, University of California Medical Center)

agent of the psittacosis-lymphogranuloma venereum group. Ornithosis, an equivalent term, defines infection by the same virus in extrapsittacine birds. Parrots, parakeets, turkeys, ducks, and pigeons are the important sources. Inhalation of dust with infected droppings or handling of feathers, viscera, or waste contaminated by infected birds may cause pneumonitis. See LYMPHOGRANULOMA VENEREUM; PNEUMONITIS; VIRUS.

The human illness, usually respiratory, may range from fatal to subclinical. While the clinical picture, results of x-ray and physical examination, and history of contact with birds (wild or domestic, living or dead) may justify bias for this diagnosis, its validity is proved only by isolation of the virus from blood or sputum, demonstration of a rise in such serological tests as complement-fixing, agglutinating, or neutralizing antibodies during convalescence, and finding the agent or other evidence of infection in birds in the patient's environment. Treatment with aureomycin (chlortetracycline) or tetracycline has been successful in man and parakeets. See CHLORTETRACYCLINE; SEROLOGY; TETRACYCLINE.

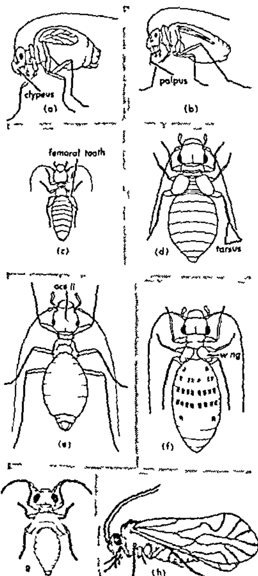
It is a fairly common occupational infection among bird breeders and dealers. The illnesses contracted from pigeons have usually been mild. Serious outbreaks have been observed among persons processing turkeys grossly diseased with a virulent strain. Cooking quickly destroys the agent. Laboratory infections are fairly common, and nurses and physicians can contract the disease from patients.

Latent infection in birds makes the disease very difficult to control. It is fairly prevalent in aviaries in the United States, so that restriction of importation of exotic cage birds only reduces the incidence. See INFECTION; RESPIRATORY SYSTEM DISORDERS. [K.F.M.E.]

Psocoptera

An order of insects frequently referred to as the Corrodentia or Copeognatha. Common names for members of this order are book lice, bark lice, and psocids. They are usually less than ¼ in. long, though rarely some may reach about ½ in. Wings are present or absent. When present, they are of differing distinctive venational types. Tarsi are 2- or 3-segmented, cerci are absent, and metamorphosis is gradual. Chewing mouthparts usually have a much enlarged clypeus, the lacinia of the maxilla is usually elongate and chisellike, and the antennae have 13 or more segments.

Book lice are most common among old papers on dusty shelves, in cereals, or other domestic situations. They are usually pale, wingless types of insects. Many bark lice, the majority winged, occur on the bark or foliage of trees, and some are found under dead bark or beneath stones. Nymphs of a few species occur on tree trunks as clusters of gregarious individuals, but disperse when mature. Food consists of microscopic molds, cereals, fragments of dead insects, and other organic debris. Sometimes psocids infest granaries, houses, straw packing, or museum collections of insects to the ex



Some genera of Psocoptera (a) *Psocathropos* (b) *Dorypteryx* (c) *L posceli s* (d) *Lep notus* (e) *Psyll psocus* short w nged adult) (f) *Trog um* (from *A B Gurney Pest Control Technology Nat onal Pest Control Assoc*) *g Lachesilla* nymph (h) *Laches lla* adult (after Sommerman)

tent that injury or a nuisance is involved but control measures are seldom necessary

Psocoptera are worldwide especially in warm countries and some 1250 species are known Current classification now lists about 26 families for this group About 150 species, in 11 families have been found in the United States and new species are being discovered frequently *See* INSECTA

[A B CU]

Bibliography D J Borror and D M DeLong *An Introduction to the Study of Insects* 1954 P J Chapman Corrodentia of the United States of America I Suborder Isotecnomera *J N Y Entomol Soc.*, 38 219 290 319–403, 1930 E L Mockford and A B Gurney A review of the psocids or booklice and barklice of Texas *J Wash Acad Sci* 46 353–368 1956

Psoriasis

An inflammatory skin disease of unknown cause usually chronic or recurrent but often acute in nature. It accounts for about 5% of all skin disorders and is most often encountered in young and middle aged adults

The local lesions occur predominantly at certain sites such as elbows knees and scalp They consist of d ll red ... ll def ...

lesions spread by peripheral extension and may involve huge areas of the body The plaques are inconstant in size shape and location during the course of the disease but they have a tendency to recur General health is not often impaired but morale may deteriorate because of the unpredictability of response to treatment Tension fatigue and overtreatment often produce flareups

The nails may become involved independently of the skin changes and show either a pitting of the surface or the formation of a brownish soft discoloration with irregular thickening Psoriasis is occasionally accompanied by a form of arthritis similar to rheumatoid Any joints are susceptible but those of the finge ...

... during flareups of the skin lesions

Each case of psoriasis requires almost individual evaluation and treatment including assessment of psychic factors *See* SKIN SKIN DISORDERS

[E C ST]

Psychic energizer

A class of drugs currently being developed for the treatment of psychotic depression and depressive response in chronic disabling illnesses Hydrazides of nicotinic acid one of the B vitamins were synthesized in 1951 for the treatment of tuberculosis By 1953 one of these iproniazid (Marsalid) was almost discarded because of its excessive stimulatory effects and the occurrence of nightmares and depression following abrupt withdrawal In 1955 it was noted that Marsalid in lower dosage induced weight gain and increased well being and activity in 1957 psychiatrists reported beneficial effects in depressed patients previously refractory to other therapies Fatal liver damage occurring in some patients led to warranted caution in use of the drug Knowledge gained concerning its inhibition of enzymes such as amine oxidase has led to synthesis of related compounds Amine oxidase destroys amines such as serotonin and adrenalin that are thought to influence brain excitability The compounds related to Marsalid are screened for the duration and potency of their amine oxidase inhibition and for behavioral effects in animals prior to clinical tests in humans *See* ENZYME SEROTONIN

Psychiatrists as yet have not disentangled the factors leading to depression with its intense gloom self persecutory thoughts motor retardation and suicidal behavior Involutional melancholias of middle age considered by some to involve endocrine factors have been long known to show moderate improvement in response to electroshock and to well known stimulants such as amphetamine or benzedrine Such stimulants and their newer derivatives such as methedrine are complex in their effects They may briefly arouse a depressed patient from his retardation but if given intravenously to a tense and excited person they may induce pleasant relaxation and unconcern and even addiction They alleviate or help one to ignore fatigue but do not improve performance in the normal person The sense of well being and improved performance although temporarily useful to a depressed person is illusory and unwarranted in the normal person The capacity for euphoriants such as cortisone morphine and cocaine to lead to toxic psychoses when continued for excess dosage requires investigation In general the effect of stimulants and energizers is dependent upon the prior condition of the individual *See* PSYCHOSIS

When given to monkeys Marsalid induces more activity during waking periods but more and deeper sleep in sleeping periods The inference that energizers induce or supply energy rather than supress pathological inhibitions to normal activity is not yet valid nor is it clear that they remove the underlying disease process In striking at symptoms which may come from various causes Marsalid may permit reestablishment of normal function by providing a period of relief from some symptoms of depression The biochemical processes through which the drug acts provide a basis for the development of future drugs and for the study of the biochemical aspects of depression

Psychiatrists will have to make a finer differentiation of the various depressed anergic and withdrawn states before it will be known how such drugs affect disease It is known that the experience of normal emotions is man s link to civilized relationships The capacity to experience love grief and mourning is necessary for reliable psychological strength and health If energizers were used to blunt such genuine personal experience instead of pathological distortions of these feelings legal control such as that used with narcotics could be expected Psychotic depression occurs with such sustained cruelty towards the self that agents ameliorating this without damaging body or brain are of immense value Current trends indicate that of all therapeutic areas in psychopharmacology the most rapid advances will be in antidepressant agents *See* PSYCHOPHARMACOLOGIC DRUGS [D X F]

Psychoacoustics

A term applied to studies of contacts between the mind and the world of sound It properly includes the production of speech as well as all aspects of hearing

Studies of speech These are made in a variety of ways The different sounds of speech may be recognized and classified and their combinations studied from the standpoint of meaning The sounds may be analyzed and measured in terms of the physical quantities frequency and acoustic pressure and the time variations of these quantities A third type of study would concern the use of vocal cords tongue lips etc in producing the different sounds For fuller treatment of some of these lines of study *see* PHONETICS SPEECH

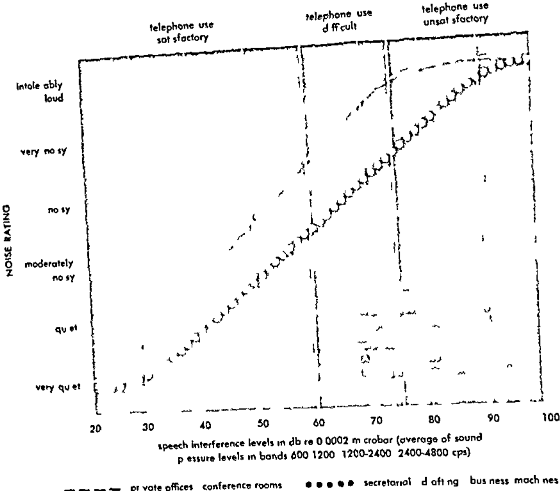
Studies of hearing Several aspects of the sensation of hearing can be connected wth the pressure and frequency of applied sounds Perhaps the most fundamental of these aspects is the smallest sound pressure that can be heard called the threshold of audibility This threshold can be measured over a wide range of frequencies Pressures of sounds within this range can then be expressed in decibels above threshold On the subjective side judgments of loudness pitch and quality are made Subjective scales of loudness and pitch have been worked out (units are the sone and the mel respectively) and have been connected experimentally with the pressure and frequency of the external sounds Quality or timbre depends upon the loudness and pitch of the various components of a complex sound

Measurements of hearing have practical applications A hearing aid for example may be tailored to an individual s use In the design of radios phonographs and other sound reproducing systems it is not always best to have every frequency reproduced with the same pressure ratio to other frequencies that it had in the original sound Loudness studies have shown that if music is to be reproduced in the home at much lower levels than the original and is to sound musically balanced low frequencies must be reproduced at somewhat higher levels than middle frequencies

For more detailed accounts of hearing studies *see* HEARING *see also* AUDIOMETRY DEAFNESS FLETCHER MUNSON CONTOURS HEARING AID LOUDNESS PITCH

Articulation testing This name has been given to a method widely used in psychoacoustics for measuring the intelligibility of speech under a given set of conditions Both speech and hearing are involved It is a direct method in which lists of syllables or words are spoken and the sounds heard are recorded by groups of listeners The percentage of sounds heard correctly is an indication of speech intelligibility Conditions which affect the scores include the intensity of the speech at the listening location the intensity of interfering noise and the amount of distortion present in the system under test which may be either acoustic or electric Articulation testing was first devised for evaluating telephone instruments and circuits but was later expanded for use in many fields The testing can extend to the effectiveness of speakers and listeners measuring defects and evaluating aids in either case

Various types of material may be used by the speakers One test uses syllables formed arbitrarily



Curves plotted from no se rat ngs g ven by employees in off ces in terms of speech interference levels (From L Beranek Acoust cs, McGraw-H ll 1954)

from d fferent consonants and vowels and having no meaning (such as mav theb etc) This method gives val ial le data on the treatment of the separate sounds of the language A quicker over all test of speech in a system is obtained from lists of monosyllal c words (such as crash, fern etc) each list of 50 words being phonetically balanced (PB) Otl er tests use spondee words (words with two equally accented syllables such as doorstep rail road etc) or complete sentences In the latter case the sc re may be based on the correct identification of certain key words in the sentence or on a response from the listener which shows he understood the sentence

Noise control criteria These have been estab lished for certain cond tions by psychoacoustic tests One imp rtant noise level is that above which damage to hearing may result For daily exposure over l ng periods of time it is believed that there is a chance f damage if the continuous noise level is m re than ab ut 85 decibels (db) above the ref erence level of 0 0002 dynes cm² or microbars (somewhat h gher bel w 300 cps) as measured in any one critical band. Brief noises may be higher perhaps 120 150 db above reference without causing damage depen ing upon the duration repetition and character of the noise and the particu lar ear involved For an explanation of the concept of critical bands *see* MASKING (SOUND) *see also* NOISE CONTROL

Speech interference criteria Below the damage level there are various stages of annoyance one of the most important being connected with interfer ence to understanding speech An example of the establishment of criteria in this respect is given in the figure The abscissa in the figure represents levels of noise found by measuring sound pressure levels (in decibels relative to 0 0002 dynes/cm²) in the three 1 octave bands between 600 and 4800 cps and averaging the three together The ordinate gives a subjective scale of noisiness ranging from very quiet through noisy to intolerably loud The curves were established by setting up different levels of noise in the rooms and questioning employees to find [illegible]

[illegible] ference rooms The letters *A* and *B* on the curves mark the upper limits of noise level which permit intelligible speech discussion The range of telephone use in the rooms is shown at the top of the figure

Levels which permit speech communication in other types of rooms, such as homes schools, theaters etc, are not quite the same Curves showing the permissible levels in octave bands covering the range up to 10,000 cps, have been worked out for some 19 different kinds of rooms

Criteria for residential areas Annoyance criteria also have been established for residential areas near factories or other noisy installations The annoyance scale ranges from "no annoyance" through "complaints" to "vigorous legal action" The noise levels associated with the various annoyance levels depend upon several facets of the disturbing noise its spectrum, whether it is continuous or impulsive, its repetitive character, and the time of day or night at which it occurs, as well as on the general background of noise in the neighborhood and the previous exposure of the neighborhood to noise [H K D]

Bibliography L. L Beranek, *Acoustics*, 1954, I J Hirsh, *The Measurement of Hearing*, 1952, S S Stevens (ed), *Handbook of Experimental Psychology*, 1951

Psychoanalysis

A term that refers to (1) a theory, (2) a method, (3) a science dealing with observation and communication in psychological disturbances and (4) a professional movement based on such a science

The science of psychoanalysis is discussed under three points of view dynamic, topographic, and economic

Dynamic theory. All mental processes may be considered as an interplay of forces Forces are originally instincts and have, according to Sigmund Freud an organic origin They are inclined to recur (repetition compulsion) with blind force and their representative ideas are emotionally charged Freud assumed that there are two large classes of instincts (1) the erotic instinct striving for life, love and union, and (2) the death instinct striving for death destruction and separation Instincts operate according to the pleasure pain principle (*Lust Unlust*, in German)

Topographic theory Freud also spoke of ego instincts which are directed to self preservation He divided the mental apparatus into functional entities first the id or reservoir of instincts or impulses and second the ego, dealing with adaptations to the external world and controlling the id instinct The ego is partly conscious and partly unconscious The unconscious part refers mostly to the so called defense mechanism controlling and counteracting the id The most important defense against painful and dangerous impulses consists either of repressing or pushing the psychological content out of the consciousness or of not admitting them to consciousness

Economic theory Psychoanalysis assumes that instincts have a definite energy and that an organism attempts to regulate the flow and discharge of energies in order to keep excitation at the lowest possible level

Freud and his disciples arrived at these theories through observations by the psychoanalytic method The psychoanalytic method consists of a total observation of the patient and his relationship with the therapist In this relationship transference and resistance are analyzed Transference means a repetition of feelings and thoughts the patient once held towards parents and other key figures and that he now holds towards the analyst. Resistance is an expression of avoidance in bringing data to the level of the conscious The manifestations of resistance are the attempts to disguise unconscious conflicts by such behavior as dreams slips of the tongue symbolic actions and uncooperative attitudes The tools of psychoanalysis are the analysis of symptoms, parapraxias dreams, and symbolism This uses the method of free association which replaces the ordinary association of logical and orderly speech by a looser form of sequences leading to an extension of awareness The role of the analyst is to develop a nonpunitive relationship of confidence with his patient to help him with the interpretation of the complex processes under analytic investigation

The main application of psychoanalytic therapy is the treatment of the psychoneuroses Modified techniques, with various degrees of success have been applied in the treatment of psychoses and psychopathic personalities Most psychoanalytic treatments are lengthy and costly Suitability for psychoanalysis of individual patients is limited and has to be determined carefully *See* NEUROSIS, PSYCHOSIS

Many modifications of Freud's classic psychoanalysis were introduced by C G Jung A Adler K Horney, O Rank E Fromm H S Sullivan and others The majority of psychoanalysts in the United States are psychiatrists although Freud recommended the training of lay analysts Analysts are organized into a number of societies according to their theoretical and therapeutic viewpoints The largest and best organized group is the International Psychoanalytic Association which Freud founded The great significance of psychoanalysis lies not merely in its importance as a treatment but also in its impact upon the behavioral and social sciences on education and art and on Western culture in general [F C R]

Bibliography C Brenner *An Elementary Textbook of Psychoanalysis* 1957 S Freud *General Introduction to Psychoanalysis* 1920

Psychogalvanic response

An electrical change that takes place in the skin of an animal or person in response to a novel or arousing stimulus The responses are detectable in those parts of the skin in which sweat glands are numerous They depend upon the sympathetic nervous system which regulates the activity of the sweat glands and the muscles of the blood vessels in the skin the sympathetic nervous system in turn being dependent upon the central nervous system and its main organ the brain However stimulation

of the sympathetic system alone suffices to elicit the electrical response

Measurement methods Two methods have been used to measure psychogalvanic responses Both require that the two electrodes placed on the skin as part of the measuring equipment must not be batteries themselves particularly not unstable ones, lest electrode artifacts be mistaken for psychogalvanic responses In one method that of J Tarchanoff one of the measuring electrodes is on an electrically active portion of the skin the other being on an inactive portion or under the skin The electrical potential difference between the two electrodes must first be reduced to zero, the psychogalvanic response is a rapid change from this condition In the second method, that of C. Fere a flow of current between the two electrodes is maintained or increased and rapid changes in this current are regarded as the psychogalvanic response Such changes are usually interpreted as changes in the resistance of the skin although the resistance of the skin is not a single value but varies with the current flowing through it The resistance of the skin appears to be due to body fluids present in the relatively dried-out outer layers of the skin The changes in resistance noted in the Fére method are probably due to potential changes in the active elements such as sweat glands as well as to the resistive changes attributable to increased amounts of body fluids The changes detected by the Tarchanoff method are due primarily to the active elements.

Stimuli The stimuli which evoke psychogalvanic responses in the intact organism are primarily stimuli which have an arousal effect that is stimuli which prepare the organism to act Such stimuli can be defined in terms of the electrical activity of the brain They are stimuli which convert the electroencephalographic activity of the brain of the relaxed waking individual into the pattern characteristic of arousal In general all stimuli when first presented have this property but only a restricted number can do so upon repeated presentations and even these produce no change if the organism is not permitted to relax In general the class of stimuli to which the organism adapts most slowly are emotional stimuli and for this reason the psychogalvanic response has been primarily regarded as an indicator of emotions It has also been used to detect lying Its success in this connection is attributable not to any intrinsic connection between lying and the psychogalvanic response but to the fact that circumstances connected with the event under investigation are particularly arousing to one person and not to another See ELECTROENCEPHALOGRAPHY LIE DETECTOR, SYMPATHETIC NERVOUS SYSTEM [J F F]

Psychology, physiological and experimental

Psychology is the science of behavior It studies all forms of behavior both human and animal from movements of one-celled animals to complex mental verbal, and social activities that are strictly human Hence it is far broader than the few topics such as intelligence testing and mental health that receive so much public attention Scientific psychology, however, confines itself to the phenomena of behavior that can be systematically and objectively observed

Psychology bridges the gap between the biological and social sciences, and it itself is both As a

cational psychology, and social psychology As a biological science it is concerned with understanding human and animal behavior as natural phenomena of living organisms, its major fields are experimental psychology, physiological psychology, and comparative psychology The articles on psychology in this encyclopedia stress the biological aspects of psychology

Experimental psychology gets its name from the use of the experimental method first introduced into psychology about 1880 This method unlike the historical and survey methods of the social sciences permits the scientist to mal

... although it cannot be used where it will interfere with normal human activities In general it is best suited to the study of the senses motor functions physiological motives, such as hunger and thirst, learning, and memory For that reason these are the topics that are encompassed by the term experimental psychology The term has come to mean not only a method but also the body of facts and principles obtained through the use of the method

The term physiological psychology refers to the study of the physiological mechanisms or correlates or behavior It attempts to relate events in the brain glands and other organs of the body to the phenomena of behavior It is therefore an interdisciplinary field in which psychologists and physiologists both work Comparative psychology is the study of phylogenetic differences and similarities in behavior among ...

experimental psychology as can much of physiological psychology Taken together however, experimental and physiological psychology cover the subject matter of psychology included in the biological or life sciences See ABNORMAL BEHAVIOR, BEHAVIOR ONTOGENY OF, BEHAVIOR AND HEREDITY, EFFECTOR SYSTEMS, EXTRASENSORY PERCEPTION (ESP), HYPNOSIS, INSTINCTIVE BEHAVIOR, INTELLIGENCE, LEARNING THEORIES, MOTIVATION, NEUROPHYSIOLOGY, PERCEPTION, PERSONALITY THEORY, REFLEX, UNCONDITIONED, SENSATION [C T M]

Bibliography C. T Morgan and E Stellar *Physiological Psychology* 2d ed, 1950, S S Stevens

Handbook of Experimental Psychology, 1951, R S Woodworth and H Schlosberg, *Experimental Psychology,* rev ed, 1954

Psychopharmacologic drugs

Drugs are natural and synthetic chemicals which affect body processes in health and disease Psychopharmacologic drugs affect body processes but *in addition significantly influence behavior, the way* a person thinks, feels, and acts Psychopharmacologic drugs were first discovered accidentally through incidental behavioral effects noted in drug treatment of physical disease They then were developed through chemical rearrangements of the parent compound and were tested on animals Psychopharmacology, a science for the systematic study and synthesis of such drugs was stimulated by the appearance of new drugs their application to mental disorders new laboratory methods, and extensive governmental financial support It is a field of wide scope Experts in biochemistry, physiology psychology, and psychiatry contribute personnel experience and investigations to psychopharmacology *See* BIOCHEMISTRY, PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL

Use Psychopharmacologic drugs are used to treat mental and physical disorders by influencing behavioral excesses and imbalances They are tools for research into social and biological processes which influence the organization of behavior In man research is applied to normal or disturbed behavior, its subjective aspects (feeling thought states of consciousness), objective aspects (sensorimotor and intellectual performance), and theoretical aspects (distribution of psychic energies, personality structure, and so on) Research on animals deals with drug effects on learning maturation drive (fear hunger, sex, punishment reward), adaptation to stress and stimuli motor activities, and associated neurochemistry

Biochemical aspects. Some new drugs resemble certain body chemicals which in unexplained ways affect mood and behavior The trail may lead from synthetic chemicals and knowledge of how they induce or improve mental dysfunction to the discovery in the mentally ill of specific aberrations in body chemistry and to discrete curative drugs Some progress has been made Failure of metabolic *systems in the breakdown of phenylpyruvic acid in* a few retarded children can be corrected by dietary adjustment Thyroid deficiency produces cretins and thyroid excess a toxic psychosis, adjustment of thyroid levels aids both *See* PHENYLPYRUVIC OLIGOPHRENIA

For the overwhelming majority of psychotic patients no specific chemical defect is known It is possible that a number of different and interacting causes and differential drug responses may lead to delineation of distinct subgroups of mental illness For 800,000 hospitalized mental patients, good social psychological and drug treatment alone may still be insufficient to restore or sustain function Therefore, research continues *on effects of* sedatives, brain surgery, coma, and convulsive procedures which have less specific and reversible biochemical and behavioral effects than do some of the newer drugs Studies proceed on the brain and behavioral effects of hormones, alcohol, opiates and essential nutrients Drugs influencing excesses of tension are studied in combination with psychosomatic problems where effects of psychic stress *may damage tissues*

Effects of drugs. No drug is limited to effects upon behavior Every drug acts through a number of body cells and cell systems Every drug acts essentially to replace, inhibit facilitate, or compete with cell chemicals which normally regulate body biochemistry Mechanism of drug action is understood in these terms No behavior is entirely conditioned by drug action Mechanisms of drug effect require consideration of cell actions and interactions with other cell systems their status prior to administration of the drug, and the character of the current environmental situation with its pattern of stimuli and opportunities for response Conditions influencing drug effects are always analyzed whether the effect be in heart muscle or in a mental attitude of cheerfulness Rather than assign a single effect to a single drug, the regularities and variabilities are accounted for by describing patterns of drug effect The ascription of energizer is simply convenient, it describes the desired but not the inevitable nor the only effect The barbiturate sedative amytal, may be sufficiently relaxing for a person to recall painful memories, but it does not compel him to tell the truth as the sobriquet "truth serum" implies

Factors influencing the effects of drugs are understood by analysis of personality, the situation and the brain Organized in agonist antagonist subsystems, the brain contains cells which retard and cells which facilitate motor activity A drug may inhibit or excite either system thus altering the balance Thus the same behavioral effect may be achieved by exciting or inhibiting opposing systems Paradoxical drug effects may depend on an unusual prior balance of these opposing systems, for example, some excited patients sleep when given stimulants and some children show excitement in response to sedatives The prior state determines the intensity of drug effect as well as some dose effects An example of the former is stimulants alerting the fatigued rather than the normal person An example of the latter is excited patients requiring sedation with amytal in a dose almost lethal to normal individuals The rate and route of drug administration influences its effect Given intravenously and rapidly, one tranquilizer induces convulsions but given by slower intravenous injection or by mouth it calms the patient Amytal sometimes produces in congenial people a cheerful intoxication instead of sleep

Study methods. The theories ascribing mental disease to abnormalities in brain chemistry are still in the unproven stage although research on brain and behavior is promising All psychopharma

cologic drugs are studied in animals for effects upon brain function through the implantation of *electrodes* for *electrically stimulating and recording* or for local drug injections Electrodes contacting single cells or at the juncture of two cells *where body chemicals in minute quantity are excreted* for an electrical impulse to pass define the way a simple environmental stimulus affects the brain and the way a drug affects the passage and fate of the stimulus Records are made of animals learning a task and the effect of drugs upon both performance and brain activity is studied specially trained animals are used in screening tests of new compounds Pharmacologists study drug effects on the transport formation and destruction of chemicals normally regulating the ability of cells to generate or inhibit impulses A core of cells *(reticular system)* is known to govern the readiness of both cortical and motor cells to discharge and hence influences behavioral sedation and alertness drugs with such effects act on components of this core Drugs clearly play a part in analysis of the processes that organize behavior *See* PSYCHIC ENERGIZER PSYCHOTOMIMETIC DRUG TRANQUILIZER

[D X F]

Psychophysical methods

Methods for the quantitative study of the relations between physical stimulus magnitudes and the corresponding magnitudes of sensation for example between the physical intensity of a light and its perceived brightness or the concentration of a sugar solution and its observed sweetness To establish these relations measurement scales are needed not only for physical magnitudes but also for subjective magnitudes Subjective scales are of course not obtained directly from observation They are theoretical models which summarize observed relations between stimuli and responses

The original three psychophysical methods first systematically worked out and published by G Fechner in 1860 were designed to yield the data from which to construct subjective scales Other techniques more direct than Fechner s are superseding his approach to psychophysical problems his methods and variants of them are still in common use however

The term psychophysical methods is sometimes extended to include certain scaling techniques which are most often used with although not limited to subjective dimensions to which there correspond no simple physical dimensions for example food preferences These techniques have in common the fact that the rationale for scaling involves the amount of dispersion of repeated or multiple judgments Detailed discussion of the method of paired comparisons method of successive intervals and other scaling techniques based on dispersion methods can be found in the bibliography at the end of this article

The remainder of this article surveys the principal methods employed in relating subjective magnitudes to physical stimulus magnitudes The two references may be consulted for greater detail regarding variants of the methods procedural details and precautions necessary to minimize biases of various kinds

Classical methods One requirement in setting up subjective scales is a method of determining subjective equality For instance one may need to equate two stimulus increments for subjective

error described below is designed for such uses

Another requirement in subjective scaling is a means for establishing the unit and the zero point of the scale Fechner's approach to accomplishing this required as a first step the measurement of two sensory quantities the absolute threshold which is the minimum stimulus energy an organism can detect and the differential threshold the minimum detectable change in a stimulus In practice both quantities must be defined as statistical averages since repeated measurement of thresholds under externally constant conditions yields moment to moment fluctuations in the values obtained In one form or another the two methods of threshold measurement devised by Fechner are still frequently used They are the method of limits (also called the method of minimal changes) and the method of constant stimuli

Method of average error This method has its principal use in the equation of subjective magnitudes Usually the subject himself adjusts a comparison stimulus to match the standard stimulus The average of a number of such settings gives the point of subjective equality and the difference between this point and the standard stimulus is the average error Two illustrative uses of the method are the measurement of accuracy of distance perception and measurement of the magnitude of so called optical illusions There are many other applications of the technique

Method of limits To measure the absolute threshold by this method the experimenter begins with a stimulus which is too weak for the subject to detect In successive [illegible]
intensity [illegible]
ject repo[illegible]
perceived [illegible] of this ascending series is reached when the subject reports that he has detected the stimulus The descending series is then begun the stimulus intensity beginning at an above threshold value and decreasing in steps until the subject signals the disappearance of the stimulus Many such series are given To help the subject avoid basing his judgments on the number of intensive steps occurring in a series the [illegible]

[illegible] from descending series the usual practice is to average the two

In measuring the difference threshold essentially the same procedure is involved except that the sub-

ject now signals the relation of a comparison stimu lus to a standard stimulus The comparison stim ulus begins at a value clearly less than that of the standard and increases in steps until the subject reports the apparent equality of the two This gives a lower difference threshold The comparison stimu lus is further increased until the subject first re ports it as greater than the standard This is the upper difference threshold The comparable proce dure is followed in a descending series which yields a descending upper and lower threshold After a large number of such trials the average of each of these four threshold values is computed The differences of each of these four averages from the value of the standard stimulus gives four values for the difference threshold Usually these four values are averaged

The most common modifications of the method of limits involve the omission of descending series in order to minimize sensory adaptation and the use of continuous rather than discrete stimulus var iation Another procedural modification is the up and down method The stimulus intensity (or inten sity difference) is decreased on the following trial if the subject has detected it increased on the fol lowing trial if he has not It is a more efficient method because it concentrates the testing in the region of the threshold although there is the dis advantage that the level of the next following stim ulation is predictable by the subject The method has proved useful however notably as shown by its use with the Bekesy audiometer for the testing of human hearing and in the study of the visual sen sitivity of pigeons *See* SENSATION

The method of limits has the virtue that it pro vides direct assessment of thresholds and is effi cient The fact that the sequence of stimulation is predictable may help the subject in concentrating but requires that care be exercised to prevent this knowledge from influencing his responses through anticipation This predictable sequence of stimula tion does not occur in the next psychophysical method to be described

Method of constant stimuli To measure the ab solute threshold the experimenter selects a small number of stimulus values in the neighborhood of the absolute threshold (previously roughly located by informal use of the method of limits) and pre- sents them to the subject a large number of times each in a prearranged order unknown to the sub ject Each time a stimulus is presented the subject reports the presence or absence of sensation

At the end of the experiment the data provide the proportion of times each stimulus was reported present by the subject One now computes the ab solute threshold as the stimulus value which has a probability of 50 of being reported present There is considerable literature concerning methods of curve fitting and interpolation for estimating this 50 point

The difference threshold is obtained from com parative judgments A standard stimulus is paired many times with each of several values of the com parison stimulus The subject reports whether the value of the comparison stimulus appeared greater or less than that of the standard The subject would sometimes prefer to say equal or doubtful using three categories of report instead of two For merly there was much debate on the question of allowing the third category the difficulty is that the values for the difference threshold are affected by the number of such doubtful judgments This in turn is influenced by the cautiousness of the sub ject The current procedure is to use two categories of report and if the subject sometimes insists on using the third the trials on which he did so are repeated later

The data resulting from the experiment are the proportions of judgments of greater as a function of the stimulus increment To obtain the difference threshold the function may be interpreted as a cu mulative distribution and the difference threshold taken as the standard deviation of that distribution. A frequent procedure is to calculate the stimulus increment for which the judgment of greater has a probability of occurrence of 75 a value halfway between a probability of 1 00 and the probability of 50 which should occur when the comparison stimu lus is equal in magnitude to the standard

The principal variations of the method of con stant stimuli have been designed to avoid the objec tion that the subject is asked to report the presence or absence of a stimulus or stimulus increment, which he knows is physically present He may be asked to report therefore in which one of several spatial locations the stimulus occurred or in which of several successive intervals of time the stimulus increment occurred

The values of absolute threshold and difference threshold obtained by these methods are in physical units for example an auditory absolute threshold is given in decibels a difference threshold for hue is specified by wavelengths in millimicrons It is clear that quite apart from the question of subjec tive measurement the psychophysical methods thus far described are valuable for obtaining physical measures of sensory resolving power and sensitivity

Fechner proposed to use the results of threshold measurement in developing a subjective metric He defined the difference threshold or just noticeable difference (jnd) as the subjective unit and the ab- solute threshold as the zero point of the subjective scale Thus the subjective intensity of a particular brightness of light for example would be specified when it was given as 100 jnds above threshold The subjective scale so defined is not a linear function of the physical stimulus scale since jnds though de fined as subjectively equal units are not of physi cally equal magnitude throughout the intensity scale The size of the jnd is approximately propor tional to physical stimulus intensity To the extent that this relation holds Fechner deduced that sub- jective intensity should be proportional to the log arithm of the stimulus intensity

There has been considerable debate over the merits of this formulation and the assumptions in volved The chief modern objection is an empirical

one that the results obtained by Fechner s methods are different from those resulting from other more direct techniques A tone 40 jnds above the thresh ... jnds ... ploy ... thods

outlined below evidence is accumulating that subjective magnitude is a power function of stimulus intensity rather than the logarithmic function required by Fechner s postulates However these same methods indicate that on at least some nonintensive dimensions jnds may be subjectively equal

Direct ratio discrimination methods Rather than requiring of the subject merely either yes no or ordinal judgments these methods require him to make direct ratio discriminations For instance he may be presented with a moderately loud tone and *then required* by turning a knob to adjust the *loudness of a comparison* tone until it is half as loud or twice as loud as the first. The first case *illustrates* the method of fractionation the second the method of multiplication In the method of magnitude estimation the subject is given a stimulus such as the brightness of a light, to serve as a modulus with a value *assigned* to it for example 10 His task as other lights of different intensities are presented to him is to assign them numbers which shall stand in the same ratio to 10 as their brightness stands to that of the modulus One twice as bright is given the designation 20 one half as bright is 5 In these and other similar methods whether the subject s task is to estimate or to produce the prescribed ratio or the prescribed fraction there are certain common characteristics Direct ratio assessments are obtained from the subject there can be experimental checks on internal consistency of the results and since the individual judgments are not of high precision repetition is required if stable averages are to be obtained

The empirical results obtained by the various methods are in fairly good agreement They agree in that to at least a first approximation subjective ...

... of subjective intensity of alternating current applied to the skin

The methods are applicable to nonintensive or qualitative attributes such as auditory pitch and hue of colors and their application is not necessarily limited to stimuli from simple physical dimensions

The direct ratio methods have not had the long history of investigation and refinement possessed by the classical and dispersion methods thus it is too soon to attempt a final assessment of their place among the psychophysical methods Similarly time has been relatively too short for classical techniques of threshold measurement to have felt much impact from signal detection theory an application of statistical decision theory to the problems of detecting signals in noise However many of the psychophysical problems whose investigation ante...

... the first psychological labora... mod... OLOGI... [J F H]

CAL AND EXPERIMENTAL

Bibliography J P Guilford *Psychometric Methods* 2d ed 1954 W S Torgerson *Theory and Methods of Scaling* 1958

Psychosis

A term that designates severe psychiatric disorders Psychosis refers usually to psychiatric conditions which are characterized among other symp...

...rective learning are outstanding characteristics of the term The *raison d'etre* for the term psychosis would be its different ...

... an intrapsychic unconscious conflict between instinctual and controlling *forces In practice there is no qualitative difference between psychosis and neurosis it is rather* one of degree From a forensic (legal) viewpoint the psychotic patient is not considered responsible *for* his acts or competent to *enter* legally binding agreements

The term psychotic is widely used but it is not a precise term as K Bowman and others have pointed out *There is probably* not too much need *for the term and it* could be dropped altogether unless it is retained as a designation for the class of disorders which includes schizophrenic manic depressive and toxic infectious disorders *and* disorders due to acute and chronic brain syndromes *See* PARANOIA PARANOID STATE SCHIZOPHRENIA

Psychoses which are primarily caused by pathological processes of the brain are *referred to as organic* psychoses and include psychoses due *to brain* tumors senile dementia and general paresis In schizophrenia and manic depressive psychoses no pathological findings can be discovered by *current methods* these psychoses are referred to as *functional* psychoses *See* ABNORMAL BEHAVIOR PARESIS GENERAL SENILE DEMENTIA [F C R]

Bibliography K M Bowman and M Rose A Criticism of the terms "psychosis" "psychoneurosis and neurosis" *Am J of Psychiat* 108 161–166 1951

Psychosurgery

The name applied to a fairly large number of operations on the brain such as prefrontal lobotomy leukotomy transorbital lobotomy and topectomy These have been recommended as tre...

... not ... use of psychosurgery in neurosis and depression *See* NEUROSIS SCHIZOPHRENIA

The techniques vary according to the specific procedure In most cases the *operation* consists of severing bilaterally the white projection fibers to

ject now signals the relation of a comparison stimulus to a standard stimulus The comparison stimulus begins at a value clearly less than that of the standard and increases in steps until the subject reports the apparent equality of the two This gives a lower difference threshold The comparison stimulus is further increased until the subject first reports it as greater than the standard This is the *upper difference threshold The comparable procedure* is followed in a descending series which yields a descending upper and lower threshold After a large number of such trials the average of each of these four threshold values is computed The differences of each of these four averages from the value of the standard stimulus gives four values for the difference threshold Usually these four values are averaged

The most common modifications of the method of limits involve the omission of descending series in order to minimize sensory adaptation and the use of continuous rather than discrete stimulus variation Another procedural modification is the up and down method The stimulus intensity (or intensity difference) is decreased on the following trial if the subject has detected it increased on the following trial if he has not It is a more efficient method because it concentrates the testing in the region of the threshold although there is the disadvantage that the level of the next following stimulation is predictable by the subject The method has proved useful however notably as shown by its use with the Bekesy audiometer for the testing of human hearing and in the study of the visual sensitivity of pigeons *See* SENSATION

The method of limits has the virtue that it provides direct assessment of thresholds and is efficient The fact that the sequence of stimulation is predictable may help the subject in concentrating but requires that care be exercised to prevent this knowledge from influencing his responses through anticipation This predictable sequence of stimulation does not occur in the next psychophysical method to be described

Method of constant stimuli To measure the absolute threshold the experimenter selects a small number of stimulus values in the neighborhood of the absolute threshold (previously roughly located by informal use of the method of limits) and presents them to the subject a large number of times each in a prearranged order unknown to the subject Each time a stimulus is presented the subject reports the presence or absence of sensation

At the end of the experiment the data provide the proportion of times each stimulus was reported present by the subject One now computes the absolute threshold as the stimulus value which has a probability of 50 of being reported present There is considerable literature concerning methods of curve fitting and interpolation for estimating this 50 point

The difference threshold is obtained from comparative judgments A standard stimulus is paired many times with each of several values of the comparison stimulus The subject reports whether the value of the comparison stimulus appeared greater or less than that of the standard The subject would sometimes prefer to say equal or doubtful using three categories of report instead of two Formerly there was much debate on the question of allowing the third category the difficulty is that the values for the difference threshold are affected by the number of such doubtful judgments This in *turn is influenced by the cautiousness of the subject* The current procedure is to use two categories of report and if the subject sometimes insists on using the third the trials on which he did so are repeated later

The data resulting from the experiment are the proportions of judgments of greater as a function of the stimulus increment To obtain the difference threshold the function may be interpreted as a cumulative distribution and the difference threshold taken as the standard deviation of that distribution A frequent procedure is to calculate the stimulus increment for which the judgment of greater has a probability of occurrence of 75 a value halfway between a probability of 1 00 and the probability of 50 which should occur when the comparison stimulus is equal in magnitude to the standard

The principal variations of the method of constant stimuli have been designed to avoid the objection that the subject is asked to report the presence or absence of a stimulus or stimulus increment, which he knows is physically present He may be asked to report therefore in which one of several spatial locations the stimulus occurred or in which of several successive intervals of time the stimulus increment occurred

The values of absolute threshold and difference threshold obtained by these methods are in physical units for example an auditory absolute threshold is given in decibels a difference threshold for hue is specified by wavelengths in millimicrons It is clear that quite apart from the question of subjective measurement the psychophysical methods thus far described are valuable for obtaining physical measures of sensory resolving power and sensitivity

Fechner proposed to use the results of threshold measurement in developing a subjective metric He defined the difference threshold or just noticeable difference (jnd) as the subjective unit and the absolute threshold as the zero point of the subjective scale Thus the subjective intensity of a particular brightness of light for example would be specified when it was given as 100 jnds above threshold The subjective scale so defined is not a linear function of the physical stimulus scale since jnds though defined as subjectively equal units are not of physically equal magnitude throughout the intensity scale The size of the jnd is approximately proportional to physical stimulus intensity To the extent that this relation holds Fechner deduced that subjective intensity should be proportional to the logarithm of the stimulus intensity

There has been considerable debate over the merits of this formulation and the assumptions involved The chief modern objection is an empirical

one that the results obtained by Fechner's methods are different from those resulting from other more direct techniques A tone 40 jnds above the threshold does not sound twice as loud as a tone 20 jnds above threshold Furthermore from the employment of the direct ratio discrimination methods outlined below evidence is accumulating that subjective magnitude is a power function of stimulus intensity rather than the logarithmic function required by Fechner's postulates However, these same methods indicate that on at least some nonintensive dimensions jnds may be subjectively equal

Direct ratio discrimination methods Rather than requiring of the subject merely either yes no or ordinal judgments these methods require him to make direct ratio discriminations For instance he may be presented with a moderately loud tone, and then required by turning a knob to adjust the loudness of a comparison tone until it is half as loud or twice as loud as the first The first case illustrates the method of [illegible]

[illegible] of a light to serve as a modulus with a value assigned to it for example 10 His task as other lights of different intensities are presented to him is to assign them numbers which shall stand in the same ratio to 10 as their brightness stands to that of the modulus One twice as bright is given the designation 20 one half as bright is 5 In these and other similar methods, whether the subject's task is to estimate or to produce the prescribed ratio or the prescribed fraction there are certain common characteristics Direct ratio assessments are obtained from the subject there can be experimental checks on internal consistency of the results and since the individual judgments are not of high precision repetition is required if stable averages are to be obtained

The empirical results obtained by the various methods are in fairly good agreement They agree in that to at least a first [illegible]
magn [illegible]
be pe [illegible]
tensity, the powers ranging from 0 3 for auditory loudness to 3 5 for subjective intensity of alternating current applied to the skin

The methods are applicable to nonintensive or qualitative attributes such as auditory pitch and hue of colors and their application is not necessarily limited to stimuli from simple physical dimensions.

The direct ratio methods have not had the long history of investigation and refinement possessed by the classical and dispersion methods thus it is too soon to attempt a final assessment of their place among the psychophysical methods Similarly time has been relatively too short for classical techniques of threshold measurement to have felt much impact from signal detection theory an application of statistical decision theory to the problems of detecting signals in noise However many of the psychophysical problems whose investigation antedates the founding of the first psychological laboratory still provide plenty of scientific grist for modern mills See HEARING, PSYCHOLOGY PHYSIOLOGICAL AND EXPERIMENTAL [J F H]

Bibliography J P Guilford *Psychometric Methods*, 2d ed, 1954, W S Torgerson *Theory and Methods of Scaling* 1958

Psychosis

A term that designates severe psychiatric disorders. Psychosis refers usually to psychiatric conditions which are characterized among other symptoms by serious impairment of judgment and power of willing or determining (volition) The rigidity of the symptoms and the marked incapacity for corrective learning are outstanding characteristics of the term *The raison d'etre* for the term psychosis would be its differentiation from the term neurosis Yet, such a differentiation is difficult Although Sigmund Freud stressed that psychotic patients suffer from "a break with reality," neurotics suffer from an intrapsychic, unconscious conflict between instinctual and controlling forces In practice there is no qualitative difference between psychosis and neurosis, it is rather one of degree From a forensic (legal) viewpoint the psychotic patient is not considered responsible for his acts or competent to enter legally binding agreements

The term psychotic is widely used but it is not a precise term as K Bowman and others have pointed out There [illegible]

[illegible] ders which includes schizophrenic manic depressive and toxic infectious disorders and disorders due to acute and chronic brain syndromes See PARANOIA, PARANOID STATE SCHIZOPHRENIA

Psychoses which are primarily caused by pathological processes of the brain are referred to as organic psychoses and include psychoses due to brain tumors senile dementia and general paresis In schizophrenia and manic depressive psychoses, no pathological findings can be discovered by current methods these psychoses are referred to as functional psychoses See ABNORMAL BEHAVIOR, PARESIS GENERAL, SENILE DEMENTIA [F C R]

Bibliography K M Bowman and M Rose A Criticism of the terms "psychosis" "psychoneurosis" and "neurosis," *Am J of Psychiat.*, 108 161-166 1951

Psychosurgery

The name applied to a fairly large number of operations on the brain such as prefrontal lobotomy, leukotomy, transorbital lobotomy, and topectomy These have been recommended as treatment for various psychoses (particularly schizophrenia and psychotic depression) severe neuroses and chronic painful conditions Most psychiatrists will not agree to the use of psychosurgery in neurosis and depression See NEUROSIS SCHIZOPHRENIA

The techniques vary according to the specific procedure In most cases the operation consists of severing bilaterally the white projection fibers to

the prefrontal areas (*see* BRAIN) The operation was introduced by E Moniz in 1936 and propagated in the United States particularly by M Freeman and J W Watts At present it is used sparingly and has been largely replaced by tranquilizing drugs (*see* TRANQUILIZER)

There is a wide discrepancy between reports of favorable results with 50% of patients becoming employed or engaged in household work and unfavorable results making the attainment of any significant advances doubtful A number of favorable results have been obtained where other methods have failed Such good cases are balanced by the danger of producing mental deterioration and a decrease of sensitivity or awareness of others and oneself which occurs in so many cases A very conservative attitude prevails toward the procedure at the present [F C R]

Bibliography O Diethelm *Treatment in Psychiatry* 3d ed 1955

Psychotherapy

A scientific method of treatment in which a trained therapist uses meaningful verbal and emotional communication to help persons who suffer from behavior disorders of psychological origin Attempts to influence mental disorders by such communication date back to the early days of mankind Although some of those methods are related to modern psychotherapeutic techniques a truly scientific psychotherapy is of very recent origin Psychotherapies may be divided into two fundamental types the analytic or exploratory method and the directive method

Analytic (exploratory) method This method consists of various forms of psychoanalysis and the dynamic psychotherapies Interpretation of transference plays a central role Analysis of dreams slips of the tongue symptoms of illness and free association are the main tools The aim of analytic methods is the uncovering of unconscious conflicts in an attempt to give the patient insight into his problems so that he can live more realistically Interpretation alone does not cure In order to respond to analytic treatment the patient must not be too ill and must be seriously motivated for therapy It also helps if the patient has better than average intelligence Social factors also limit the applications of analytic psychotherapy

Directive method In this method there is a parental or authoritative relationship between the therapist who knows or assumes he knows the answers to the patient s problems and the patient who is influenced by his advice suggestion hypnosis reassurance or environmental manipulation The directive psychotherapist employs techniques that are relatively simple and less time consuming than those of the analytic psychotherapist They depend in large degree on the personality of the therapist and his relationship with the patient He is more inclined to combine or supplement his psychotherapy with drugs or other organic means of treatment The directive methods seem more widely applicable but a less radical change of personality is expected The aim of directive psychotherapy is the alleviation of guilt fears and insecurity by a friendly supportive relationship

Within the analytic group there are several schools of thought differing in method and theory Schools were founded by C Jung A Adler and O Rank who disagreed with their teacher Sigmund Freud on important points Others who disagreed and have considerable following in America are K Horney E Fromm and H S Sullivan Some therapists claim independence from any school of thought using any technique which helps their patients but true eclectic psychotherapists are rare J Taft a psychiatric social worker and C Rogers a clinical psychologist have evolved important systems of psychological intervention in behavior disorders Recent experimental psychological views fused with psychoanalytic views were introduced by J Dollard N Miller and others

Few hard facts are known about success or failure in psychotherapy Symptom and character neuroses particularly hysterical reactions phobias and anxiety states have responded to psychotherapy Obsessive compulsive reactions and psychosomatic disorders are more difficult to treat although success varies from one disease category to another and from one patient to another This is true also for the difficult task of treating the functional psychoses *See* NEUROSIS OBSESSIVE COMPULSIVE REACTION PHOBIC REACTION

The shortage of trained psychotherapists creates an acute problem The long standing therapeutic tradition of the medical profession and the confidence of the public are instrumental in giving prestige to medical psychotherapists The training of nonmedical psychotherapists such as psychologists social workers and clergymen is a relatively new and less organized field Since psychotherapy is essentially a process of emotional reeducation there are some who believe that a new profession of nonmedical psychotherapists should be created

Practical problems in evaluation of treatment and development of better and more applicable techniques are important as well as a more rigorous theoretical approach which might transform an old art into both a new science and a sound technique As this gradually happens the direct and indirect impact of psychotherapy on many aspects of human life may become great *See* PSYCHOSIS

[F C R]

Bibliography F Alexander and H Ross (eds) *Dynamic Psychiatry* 1952 J Dollard and N E Miller *Personality and Psychotherapy* 1950

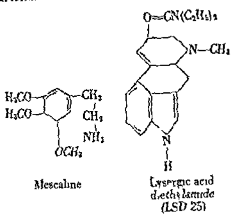
Psychotomimetic drug

A class of drugs reliably inducing temporary states of altered perception often with symptoms similar to those of psychosis Their classification emerged from revived research in the biological determinants of mental disorder Model psychoses induced by mescaline were studied in the nineteenth century Now with drugs which not only induce

and maintain but also temporarily alleviate mental

PSYCHOPHARMACOLOGIC DRUGS PSYCHOSIS

Mescaline and LSD 25 The two drugs of chief interest mescaline and LSD 25 (lysergic acid diethylamide) are of ancient lineage Their structural formulas are



Mescaline

Lysergic acid diethylamide (LSD 25)

Mescaline used in Amerindian tribal religious rituals is derived from peyote buttons which are the dried tops of a cactus found in southwestern America A Stoll and W A Hoffman prepared LSD 25 in 1938 from ergot a fungus which infests rye In this unpurified form it plagued medieval man with outbreaks of abortion gangrene and madness Along with their mental effects two features of these drugs are striking the minute quantity of LSD 25 needed to produce its effect and the similarity of both drugs to amines such as indoles adrenalin and serotonin which occur naturally in the body These features indicate that but a few molecules of a hypothetical body substance hitherto eluding detection would be necessary to produce psychotic effects and that a disorder in amine metabolism could be a source of mental disorder This is known to occur in phenylpyruvic mental deficiency See PHENYLPYRUVIC OLIGOPHRENIA The

... proposed to be body substances of interest as yet no substance from normal or mentally ill persons has been isolated and shown une... ...call ...

... for changes in brain enzymes which are common to the amines and psychotomimetics as well as for tranquilizers such as reserpine Identification of the enzymes through wh... LSD 25 ... standing ... function ... havior Continued discovery of psychotomimetics such as those from mushrooms which are linked to brain receptors for acetylcholine is spurred by the hope that their chemical analysis will be a clue to different biochemical lesions in mental disease

Relation to research on psychosis The discovery of any fundamental biochemical mechanism would in itself solve few problems In the evolution of a psychosis a biochemical lesion could be critical in inhibiting the normal development of psychological assets or inducing a break with reality or in maintaining or enhancing it or in inhibiting recovery of function Whatever the role of biochemical factors they are not even in model psychoses sole determinants of the disorder Most mental disease is a disorder in behavior and lacks any known physiologic or biochemical stigmata Considerable psychologic sophistication is required to differentiate critical phases and attributes of disordered behavior if the role of biochemical factors is to be properly understood

The description of psychotomimetic disorder its comparison with mental disease and the differentiation of psychopharmacologic drug effects is a function of psychiatry Hashish opium alcohol and cocaine observably alter relationships with reality but bring different psychic gains and results than LSD 25 LSD 25 and mescaline may change perception without any objective signs and when taken daily they induce tolerance to the drugs This may be considered as a mechanism of resistance which could provide a biochemical defense for psychoses Addiction and habituation to these drugs are not produced nor is relief from pain and anxiety Unlike LSD 25 drugs inducing toxic psychoses due to excessive dose cause a clouded sensorium and confusion some like the antimalarial atabrine induce delusions and are of investigative interest Hormones like cortisone occasionally produce psychoses

The fasting of the ascetic the exhaustion of the sleep deprived soldier the ritual trance from rhythmic dancing the isolation from stimuli and response of the prisoner in confinement and the retreat to sleep of man induce psychophysiologic states in which illusion dream and hallucination can occur LSD 25 and mescaline similarly have the property of disengaging one's usual orientation to the everyday order of things they then may have the effect of facilitating exalted visionary states or of facilitating psychoticlike states Culture situation personality and drug determine the kind of experience occurring during the disengagement Fundamentally one discerns a diminished mental control and the coexistence of primitive and logical thought and perception The usual aspects of reality are of less interest and quaint aspects are selected for attention The subject's feelings and momentary ne... the norma... omissions lacking a... ... occupies the attention becomes at the moment compellingly significant Objects seen a moment ago persevere along with current perceptions—giving rise to reports of hallucination Thinking and perceiving of this order

coexist with a capacity for, but not an interest in, normal thought and function Reactions to such shifts in experience may be pleasant, as in the ritual situations or in individual aesthetic experi ence or they may be dysphoric leading to fear, paranoid thoughts, depersonalization, and mood swing

The dependence of the symptomatology upon the presence and actions of the physician has led to the use of LSD 25 in therapy The aims and organiza tion of the subject also determine whether altered perceptions are experienced with aesthetic interest and insight or as frightful hallucinations The dose of the drug can heighten the disengagement, the oc currence of hallucination, and contentless affective outbursts, but the fundamental psychological state in which these events proceed is dependent only upon minute quantities of drug The reaction is over in 8–10 hours [D X F]

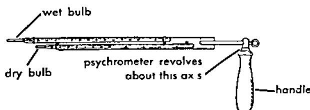
Psychrometer

A device for measuring the relative humidity (RH) or moisture content of air or gas by two temperature measurements, one dry bulb and one wet bulb Relative humidity, or per cent saturation, is usually expressed as a percentage

To obtain satisfactory readings on the two ther mometers the air or gas sample must circulate past the bulbs with a velocity of 900 ft/min or higher This ensures a good dynamic response from the dry bulb and exposes the wet surface to a true sample of the gas A fabric wick covers the wet bulb and its connecting structure, and the thermometer measures the temperature of the water (distilled) which is conveyed to it by the wick At 100%RH, or saturation, the readings on the wet and dry bulb thermometers will be identical, but the de *pression of the wet bulb reading will increase as* the moisture content of the gas is lowered (RH decreases)

At low temperatures, for example 40°F, the wet bulb depression with dry air (0%RH) is 13° and at high temperatures, say 170°F, the depression with dry air is 90°F The relationship is nonlinear, and a psychrometric chart or slide rule is used to convert the thermometer readings into humidity (pounds of water per pound of gas) relative hu midity (% saturation) or dew point (temperature at which gas becomes saturated)

The sling psychrometer illustrated combines two etched stem thermometers on a frame with a swivel mounted handle at one end The device is



Sling psychrometer (From D M Considine, ed, Process Instruments and Controls Handbook, McGraw Hill 1957)

whirled rapidly to give proper air velocity, after which readings are taken rapidly to obtain accu rate results

The Assman psychrometer utilizes a power driven fan to circulate the air sample over the two thermometer bulbs In industrial practice, this cir culation method is used with all types of thermal elements to measure the moisture content of air and gases In home and laboratories, periodic psy chrometric readings are commonly obtained by manually fanning a wall psychrometer

The sling and Assman psychrometers, when ac curately calibrated and properly used, give accu racies approaching 2% between 20 and 80%RH and 50 to 110°F Errors as great as 5%RH are possible at temperature and humidity extremes With household and industrial psychrometers, a 5% accuracy can be expected with a good installa tion *See* HYGROMETER, MOISTURE CONTENT MEAS UREMENT [R E CL]

Bibliography D M Considine (ed), *Process Instruments and Controls Handbook,* 1957, O T Zimmerman and I LaVine, *Psychrometric Tables and Charts,* 1945

Psychrometrics

A study of the physical and thermodynamic prop erties of the atmosphere The properties of pri mary concern in air conditioning are (1) dry bulb temperature, (2) wet bulb temperature (3) dew point temperature, (4) absolute humidity, (5) per cent humidity, (6) sensible heat, (7) latent heat (8) total heat, (9) density, and (10) pressure

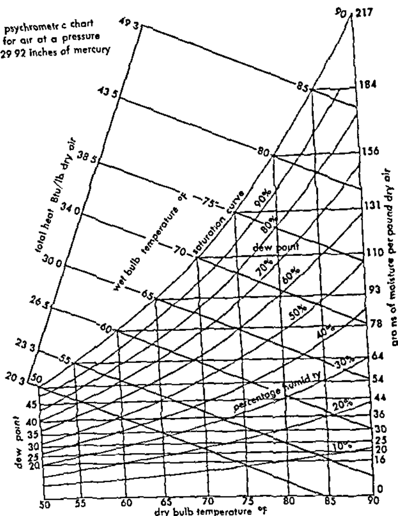
The atmosphere in a clean pure state is a me chanical mixture of dry air and water vapor Each is independent of the other and follows the laws of physics in accordance with its respective physical properties *Normal atmosphere contains impuri* ties such as carbon dioxide, ozone, dust, and dirt in varying quantities

Graphic representation Even though dry air and water vapor are independent entities there are certain relationships which permit the inclusion of these factors on a common chart This is known as a psychrometric chart as shown

The psychrometric chart is a valuable tool for understanding the relationship of the many vari ables encountered in the atmosphere and for solv ing problems in air conditioning

Dry bulb temperature The dry bulb temperature is the ambient temperature of the air and water vapor as measured by a thermometer or other tem perature-measuring device in which the thermal element is dry and shielded from radiation Tem perature scales are usually in degrees Fahrenheit or degrees Centigrade The vertical lines on the chart are dry bulb temperature lines

Wet bulb temperature If the bulb of a dry bulb thermometer is covered with a silk or cotton wick saturated with distilled water and the air is drawn over it at a velocity not less than 1000 ft/min, the resultant temperature will be the wet bulb temper ature



Psychrometric chart for air at a pressure of 29 92 inches of mercury

If the atmosphere is saturated with water vapor the water on the wick cannot evaporate and the wet bulb will read the same as the dry bulb If the atmosphere is not saturated water will evaporate from the wick at a rate dependent upon the percentage of saturation of the atmosphere The cooling produced by the evaporation will result in a lowering of the temperature of the bulb and the consequent reading is the wet bulb temperature

The lines on the psychrometric chart extending from upper left to lower right are wet bulb temperature lines Where the dry bulb and wet bulb temperatures are the same the atmosphere is saturated and a line drawn through these points is called the saturation or 100% relative humidity curve

Dew point temperature The dew point temperature is the temperature at which the water vapor in the atmosphere begins to condense This is also the temperature of saturation at which the dry bulb wet bulb and dew point temperatures are all the same

The dew point temperature is a measure of the actual water vapor content in the atmosphere The water vapor content is constant for any dew point temperature regardless of the dry bulb or wet bulb temperature The dew point temperatures are shown as the horizontal lines on the psychrometric chart

Absolute humidity The actual quantity of water vapor in the atmosphere is designated as the absolute humidity It is measured as pounds of water vapor per pound of dry air grains of water vapor per pound of dry air, or grains of water vapor per cubic foot of dry air One pound of water vapor is equal to 7000 grains Because the absolute water vapor content is directly related to dew point temperature the vapor content in grains per pound of dry air is shown on the vertical scale on the right hand side of the psychrometric chart for different dew point temperatures

Percentage or relative humidity Percentage or relative humidity is the ratio of the actual water vapor in the atmosphere to the quantity of water vapor the atmosphere could hold if it were saturated at the same temperature For example if the atmosphere at 50°F is saturated with water vapor, there would be 53 grains of water vapor per pound of dry air If the dry bulb temperature of this atmosphere were then raised to 69°F, it could hold 106 grains of water vapor (corresponding to a saturation temperature of 69°F) Because no moisture was added in changing the temperature from 53°F to 69°F the actual vapor content is still 53 grains

Table 1 Properties of water

Temperature °F	Pressure psia	Enthalpy of evaporation Btu/lb
40	0 122	1070 64
45	0 147	1067 81
50	0 178	1064 99
55	0 214	1062 16
60	0 256	1059 34
65	0 306	1056 52
70	0 363	1053 71
75	0 430	1050 89
80	0 507	1048 07
85	0 596	1045 23
90	0 698	1042 40
95	0 816	1039 56
100	0 950	1036 72

The percentage humidity would then be 53/106 or 50% at 69°F

Sensible heat Sensible heat or enthalpy of dry air is heat which manifests itself as a change in temperature It is expressed in British thermal units (Btu) One Btu will raise the temperature of 1 lb of water 1°F One Btu will also raise the temperature of 1 lb of dry air approximately 4 25°F at atmospheric temperatures and sea level pressure

Latent heat Latent heat or enthalpy of vaporization is the heat required to change a liquid into a vapor without change in temperature For example it would require 1054 Btu to change 1 lb of water at 70°F from a liquid to a dry saturated vapor at 70°F

Latent heat is sometimes referred to as the latent heat of vaporization and varies inversely as the pressure The higher the pressure (or saturation temperature) the lower the Btu required to evaporate water from a liquid to a vapor Table 1 shows the thermodynamic properties of water for various temperatures

Total heat The total heat or enthalpy of the atmosphere is the sum of the sensible heat latent heat and superheat of the vapor above the saturation or dew point temperature At saturation the total heat measured in Btu/lb dry air is measured at the wet bulb temperature and includes both the sensible heat of the dry air and the latent heat of the water vapor at the temperature measured For example at 60°F on the saturation curve the total heat is

Table 2 Temperature volume-enthalpy relation of saturated air

Temperature °F	Volume ft^3/lb of dry air	Enthalpy Btu/lb of dry air
40	12 70	15 23
45	12 85	17 65
50	13 00	20 30
55	13 16	23 22
60	13 33	27 15
65	13 50	30 06
70	13 72	34 09
75	13 88	38 61
80	14 09	43 69
85	14 31	49 43
90	14 55	[illegible] 93
95	14 80	63 3[illegible]
100	15 08	71 73

26 46 Btu/lb of dry air with water vapor to saturate the air at that temperature

Total heat is relatively constant for a constant wet bulb temperature deviating only about 1 5–2% low at relative humidities below 30% Table 2 shows enthalpy values for saturated air

Density The density of the atmosphere varies with both altitude and percentage humidity The higher the altitude the lower the density and the higher the moisture content the lower the density

At sea level (29 92 in Hg absolute pressure) and 59°F the density is 0 0765 lb/ft^3 At 5000 ft elevation and 59°F the density would be 0637 lb/ft^3

At sea level and 65°F saturated the density is 0 074 lb/ft^3 At 65° dry bulb and 30% saturation the density is 0752 lb/ft^3 The reciprocal of density (*cubic feet per pound of dry air*) is usually used rather than the density Table 2 shows temperature volume relations for saturated air

Pressure Atmospheric pressure usually referred to as barometric pressure is measured either in inches of mercury (29 92 in Hg at sea level) or in pounds per square inch absolute (14 7 psia at sea level)

Pressure varies inversely as elevation as temperature and as percentage saturation Pressure decreases with elevation as is shown in Table 3

Table 3 Altitude and pressure (standard atmosphere)

Altitude, ft	Pressure in Hg
−1 000	31 02
−500	30 47
0 (Sea level)	29 92
500	29 38
1 000	28 86
5 000	24 89
10 000	*20 58*

See COMFORT CONTROL [J F\ ]

Bibliography American Society of Heating and Air Conditioning Engineers *Heating Ventilating Air Conditioning Guide* 37th ed 1959 W H Carrier Rational psychrometric formulae *Trans Am Soc Mech Engrs* 33 1005 1053 1911 W H Carrier and C O Mackey *A Review of Existing Psychrometric Data in Relation to Practical Engineering Problems* Am Soc Mech Engrs Advance Paper 1936 J A Goff Thermodynamic properties of moist air *Heating Piping and Air Conditioning* 6(3) 117 132 1934 W K Lewis The evaporation of a liquid into a gas—a correction *Mech Eng* 55(9) 567 568 1935 D D Wile Psychrometric charts *Am Soc Heating Air Conditioning Engrs Journal* vol 1 no 8 1959

Ptarmigan

Any of four species of small grouse of the genus *Lagopus* family *Tetraonidae* The ptarmigans are brown and more or less typical grouse in the summer but they undergo a seasonal plumage change in the fall gradually changing to white for the winter season and then turning brown again in the spring Two species are found in the United



The willow ptarmigan Lagopus lagopus length to 17 n (From P M Duncan ed Cassell's Natural H story Cassell)

States The willow ptarm gan *L lagopus* is a bird of the Arctic tundra which occasionally wanders into the northern border states in winter The *wh te-tailed* ptarmigan *L le icurus* nests in the mounta ns above the tree line from Alaska south into New Mexico *See* CALLIFORMES GROUSE

[J B B]

Pteraspidomorphi

gie exhalent pore and absence of paired appendages The lateral eyes are widely separated on the broad depressed head Exoskeletal plates cover the head and the unpaired rostral is quite prolonged *Pterasp s* is a well known example *See* AGNATHA CEPHALASPIDOMORPHI COELOLEPIDA HETEROSTRACI

[C B C]

Pteridospermae

Seed ferns extinct plants characterized by naked seeds borne on large fernlike fronds Fossil evi dence indicates that these plants reached their greatest abundance dur ng the Pennsylvanian period of geological time approximately 260 000 000 years ago Subsequently they decl ned becoming extinct sometime during the Jurassic period or about 175 000 000 years ago

Becau e some seed ferns had rather long slender stems with relatively small amounts of supporting t sue it has been surmised that they were vine l ke and requ red the support of more sturdy neigh l ring plants Others grew erect without the a d of other plants and in their general appearance the

p

e u-20 ft Near the apex of the r stems they produced a crown of large sp rally arranged fronds (F g 1) Both the fronds of tree ferns and the pteridosperm leaves were com p nd and they both attained an estimated length of 3-5 ft Slender adventitious roots were b rne on the lower parts of the stem

Classification The study of pteridosperms has added to botanists knowledge of relationships among vascular plants Before their discovery and elucidation botanists generally agreed that there was little or no relationship between ferns which have no seeds and seed bearing plants When it was proved that pteridosperms with their fern like features formed seeds similar in structure to modern cycads it was concluded that ferns and cer tain seed plants (Cycadophytes) evolved from a common ancestor and thus were more closely related than was previously suspected Because of the paleobotanical evidence and evidence from comparative anatomy of extant vascular plants a taxon—the Pteropsida—was proposed by the late E C Jeffery of Harvard University to include the classes Filicineae (true ferns) Gymnospermae (cycapophytes conifers ginkgos) and Angiosper mae (flowering plants) In this classification the Pteridospermae (*see* CYCADOFILICALES) is an order of the class Gymnospermae In another classifica tion which recognizes three unrelated groups of naked seeded plants Pteridospermae is an order of the Division Cycadophyta This division also in cludes Cycadeoidales Caytoniales and Cycadales

Pteridospermae are generally divided into three families Medullosaceae Lyginopteridaceae and Calamopityaceae Of these the medullosan and lyginopterid pteridosperms are the best known

Medullosaceae Three stem genera *Medullosa Sutcliffia* and *Colpoxylon* are assigned to this fam ily *Medullosa* and *Sutcliffia* are consistently poly

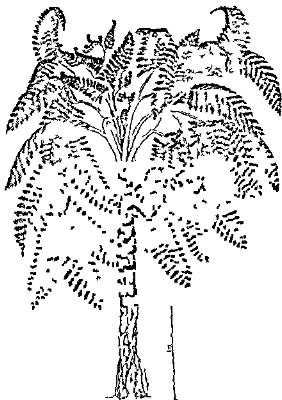


Fig 1 Reconstruction of a seed fern

stelic, *Colpoxylon* may have a single, large, irregular stele All have large spirally arranged petioles with numerous vascular bundles

Stem anatomy Depending upon the position in the stem the number of steles in a species of *Medullosa* may vary from 2 to 23 In some species the number is more consistently 2–3 Each stele is composed of a centrally located, mixed protostele consisting of mesarch primary xylem in a parenchymatous ground tissue A similar arrangement of stelar tissues is found in stems of many true ferns Around each protostele there developed an active vascular cambium that produced conspicuous secondary wood composed primarily of large tracheids with multiseriate bordered pits on their radial walls Exceptionally well preserved secondary phloem is usually found to the outside of the vascular cambium The parenchymatous fundamental tissue in which the steles are embedded is delimited externally by a layer of periderm Stem specimens near the apex have spirally arranged leaf bases containing many vascular bundles and vertically extended, peripherally disposed strands of thick walled supporting cells These leaf bases (petioles) are sloughed off from older more basal parts of the stem (Fig 2) *See* STELE

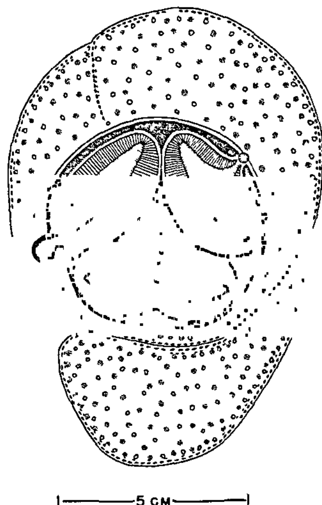


Fig 2 Cross section of *Medullosa* stem

Leaves Isolated petioles of leaves are placed in the genus *Myeloxylon* These may branch dichotomously 2–3 times Ultimate divisions of this branching system bear alternately disposed pinna Certain pinna assigned to the genera *Alethopteris* and *Neuropteris* (Fig 3) are now known to have been attached to *Myeloxylon*

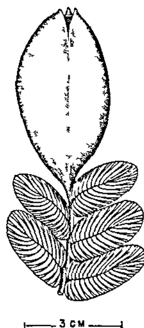


Fig 3 *Neuropteris* pinna with attached seed

Roots Only adventitious roots are known for Medullosaceae These are borne on the stem between the leaf bases Structurally mature roots usually consist of a tetrarch and exarch protostele with four segments of secondary xylem alternating with arms of primary xylem A periderm formed the outer covering of older roots

Seeds These reproductive organs are large—4–6 cm long and 1–2 cm in their greatest diameter Compression fossils show that in certain species, seeds replaced the pinnule at the tip of a pinna rachis Others apparently replaced lateral pinnules In either case the rest of the frond remained unmodified A thick vascularized testa (seed coat) formed the outer covering This prompted the generic name *Pachytesta* (thick skin) for isolated

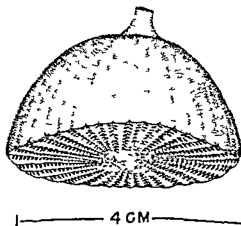


Fig 4 *Dolerotheca* sp

ds of this type The apex of the testa was di ed into three valves along three commissured s Within the testa was a single large stalked gasporang um apically differentiated into a pol n chamber Pollen grains of the *Monoletes* type ave been found in these pollen chambers and in least one instance a microgametophyte was ound Megagametophytes are usually lacking but everal instances of their occurrence have been re orted Archegonia are occasionally preserved with gg protoplasts In their structures these seeds and heir included megagametophytes are similar to seeds of modern cycads and ginkgo

Pollen forming organ Dolerotheca the pollen forming organ has the shape of a bell (campanu lum) consisting of a stalk and a vascularized cover from which arise double rows of pendant fused sporangia The sporangia are linear and form pol len of the *Monoletes* type They have never been found attached but anatomical studies show that they replaced pinna on the medullosan frond (Fig. 4)

The monolete pollen grains are relatively large some measuring over 250 microns in their greatest dimension They are bilaterally symmetrical with a pair of grooves on the distal surface and a slightly angled proximal suture In many ways they resemble pollen of some ranalian angiosperms *See* POLYNOLOGY

Lyginopteridaceae Members of this family in clude monostelic pteridosperms having one or two vascular traces entering the base of the petiole Common representatives are the stem genera *Het erangium* and *Lyginopteris* Foliage may be typi

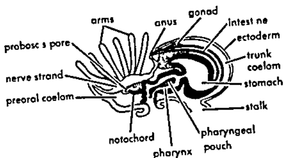
pteridosperm feature or features are known These include seeds pollen producing organs foliage stems and the like some of which are placed in the third family Calamopityaceae Although the pteridosperms were recognized in 1903 the under standing of this group did not begin until the 1950s Evidence indicates that cycadeoids and modern cycads evolved from pteridosperms Mem bers of this great extinct complex of plants may provide the answer to the perplexing question of origin of the angiosperms *See* PALEOBOTANY

[W N S]

Bibliography C A Arnold *An Introduction to Paleobotany* 1947 W N Stewart and T Delevor yas The medullosan pteridosperms *Botan Rev.*, 22 45–80 1956

Pterobranchia

A group of animals regarded as a class of the Hemi chordata They are all marine widely distributed but not common benthonic (chiefly deep-sea) in hab tants. These organisms are small or micro-



Cephalodiscus diagrammatic view (After W Patten from H V Neal and H W Rand Comparative Anatomy Blakiston Division McGraw-Hill 1936)

scopic sessile and tube dwelling (tubicolous) They may be colonial or pseudocolonial with a cu ticular skeleton A U shaped gut is present One pair of gill slits or none are present A lophophore of one or several pairs of arms arises from the col lar or middle of the three body segments The pro boscis is flattened into a preoral disc and the trunk is short but has a long contractile stalk These

Pteropsida

A subphylum of plants distinguished by their com plex water and food conducting systems large showy leaves and other advanced features which set them apart from the lower forms A subdivision of the phylum Tracheophyta the Pteropsida con sists of three classes Filicineae (ferns) Gymno- spermae (conifers and their allies) and Angio- spermae (flowering plants) *See* ANGIOSPERMAE FILICINEAE GYMNOSPERMAE *see also* PLANT KINGDOM [P D S]

Bibliography See EMBRYOPHYTA

Pterosauria

An extinct order of flying reptiles of the Mesozoic Era A thin membrane of skin extending from the side of the body to the enormously elongated fourth finger of each hand formed a wing analogous to that of a bat The first three fingers were short and clawed The small rear limbs bore pentadactyl feet All bones were extremely light and pneumatic The sternum was large The Jurassic suborder Rhamphorhynchoidea had long slender tails with



Fig 1 Restoration of *Rhamphorhynchus phyllurus* from the Late Jurassic of Germany (After O C Marsh)

stelic, *Colpoxylon* may have a single, large, irregular stele All have large spirally arranged petioles with numerous vascular bundles

Stem anatomy Depending upon the position in the stem the number of steles in a species of *Medullosa* may vary from 2 to 23 In some species the number is more consistently 2-3 Each stele is composed of a centrally located, mixed protostele consisting of mesarch primary xylem in a parenchymatous ground tissue A similar arrangement of stelar tissues is found in stems of many true ferns Around each protostele there developed an active vascular cambium that produced conspicuous secondary wood composed primarily of large tracheids with multiseriate bordered pits on their radial walls Exceptionally well preserved secondary phloem is usually found to the outside of the vascular cambium The parenchymatous fundamental tissue in which the steles are embedded is delimited externally by a layer of periderm Stem specimens near the apex have spirally arranged leaf bases containing many vascular bundles and vertically extended peripherally disposed strands of thick walled supporting cells These leaf bases (petioles) are sloughed off from older more basal parts of the stem (Fig 2) *See* STELE

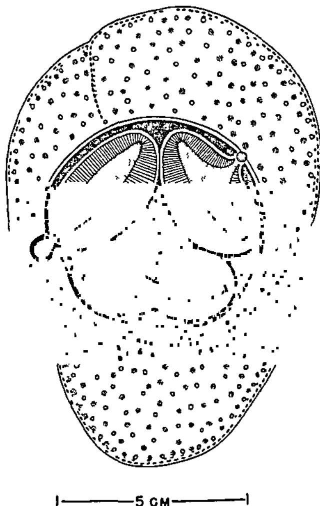


Fig 2 Cross section of *Medullosa* stem

Leaves Isolated petioles of leaves are placed in the genus *Myeloxylon* These may branch dichotomously 2-3 times Ultimate divisions of this branching system bear alternately disposed pinna Certain pinna assigned to the genera *Alethopteris* and *Neuropteris* (Fig 3) are now known to have been attached to *Myeloxylon*

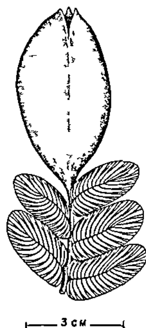


Fig 3 *Neuropteris* pinna with attached seed

Roots Only adventitious roots are known for Medullosaceae These are borne on the stem between the leaf bases Structurally mature roots usually consist of a tetrarch and exarch protostele with four segments of secondary xylem alternating with arms of primary xylem A periderm formed the outer covering of older roots

Seeds These reproductive organs are large—4-6 cm long and 1-2 cm in their greatest diameter Compression fossils show that, in certain species seeds replaced the pinnule at the tip of a pinna rachis Others apparently replaced lateral pinnules In either case the rest of the frond remained unmodified A thick vascularized testa (seed coat) formed the outer covering This prompted the generic name *Pachytesta* (thick skin) for isolated

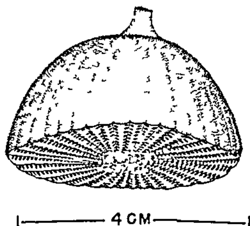


Fig 4 *Dolerotheca* sp

Fig 2 Block diagram of an elaborate institutional public address system

arks or factory areas The type of speaker used most frequently in small systems or in systems having a large number of distributed speakers is a high quality single cone speaker about 8 in in diameter such as that shown in Fig 4c

Speaker location. In a long room with a low ceiling many speakers distributed throughout the room are required For a sound reinforcing system in an orchestra hall however, loudspeakers are usually located over the center of the proscenium arch or on each side of the stage When a human speaker is addressing the audience, the best position for the loudspeaker is over the center of the stage, otherwise the sound from a loudspeaker on one side of the stage might arrive at the listener's ear from a direction different from the direction from which the listener's eyes tell him the sound should come The directional characteristics of the speakers must be chosen to distribute the sound uniformly to all angular directions where people are seated and yet direct very little energy against the walls Sufficient level must also be provided in the distant parts of the auditorium without blasting the people near the speakers Sound directed toward the walls does no good and may do harm by increasing the danger of the unpleasant resonance effect known as singing

Oscillation or singing Oscillation is a perpetual source of difficulty in sound reinforcing systems

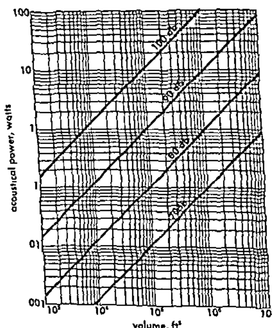


Fig 3 The acoustical power required to produce sound *intensity* levels of 70, 80, 90, and 100 decibels (db) in an auditorium plotted as a function of the volume of the auditorium (*From H F Olson, Acoustical Engineering, 3d ed, Van Nostrand, 1957*)



Fig 2 Restoration of the skeleton of the giant Cretaceous flying reptile, *Pteranodon* (After Eaton)

an expanded tip used to control flight (Fig 1) The Pterodactyloidea of the Late Jurassic and Cretaceous lacked tails Their dentition was varied some had stout, sharp, forwardly directed teeth presumably adapted for catching fish or small terrestrial animals, others with numerous thin needlelike teeth must have been insectivorous *Pteranodon* (Fig 2), the giant pterosaur of the Cretaceous, with a wing span of nearly 7 m and crested skull 84 cm long was toothless and presumably a fish eater as its remains are abundant in deposits formed far out at sea *See* ARCHOSAURIA, REPTILIA

[J T C]

Bibliography H G Seeley, *Dragons of the Air, an Account of Extinct Flying Reptiles* 1901

Pterygota

A subclass of the Insecta The majority of these insects are characterized by having wings present in the adult stage Some are wingless, as in the lice or fleas, while others have rudimentary or reduced wings Abdominal appendages, such as styli, which are fingerlike structures found commonly among the primitive, wingless insects, are lacking External genital structures and cerci may be present on the posterior abdominal segment This subclass is divided into two divisions, the Exopterygota or Hemimetabola and the Endopterygota, or Holometabola The division of this subclass into these two groups is based on the type of metamorphosis or change during development to the adult stage *See* APTERYGOTA, ENDOPTERYGOTA, EXOPTERYGOTA, INSECTA

[E O E]

Bibliography E O Essig *College Entomology*, 1942

Ptychodactiaria

An order of the zoantharian anthozoans of the phylum Coelenterata This group of solitary sea anemones is known only from two genera *Ptychodactis* and *Dactylanthus* from the Arctic and Antarctic The specialized mesenterial filaments of the actinians are absent and the mesenteries are not arranged in good cycles *See* ANTHOZOA, COELENTERATA

[C H]

Public address system

A specialized form of sound reproduction system in which sound primarily from a live sound source, is presented to a large audience by means of loudspeakers, often called simply a PA system A public address system is necessary when the area to be covered with sound is so large or the noise level so high that the original human speaker or musical group cannot generate enough sound energy to be heard throughout the area Provision is often included in public address systems for reproducing recorded or broadcast musical programs in addition to the live source of sound There are also specialized forms of PA systems, for example, in schools and hospitals intercommunication provisions are included to permit the teacher or patient in an individual room to communicate with a central point such as the principal's office or the nurse's station

PA systems vary from the simplest form shown in Fig 1 to more elaborate PA systems containing most or all of the elements of the block diagram shown in Fig 2 When no operator is present to control the gain or volume and several people may be speaking over the system the compressor amplifier in Fig 2 may be used to prevent overload of the output power amplifier and loudspeakers

The acoustical power necessary to produce the specified sound levels in auditoriums of different sizes can be calculated from Fig 3 For example in an auditorium 50 by 100 by 20 ft (volume of 100 000 ft^3) slightly more than 10 acoustic watts would be required to reproduce a 100-db level corresponding to the extreme peak sound levels of a symphony orchestra To obtain this level with single-cone direct radiator loudspeakers of moderately high efficiency say 5%, 200 watts of amplifier power would be required but with relatively high efficiency horn speakers having an efficiency on the order of 30% an electrical power of only 33 watts would be required *See* LOUDSPEAKER

PA system loudspeakers Speakers used in PA systems vary from the elaborate multicell high frequency horn speakers with large bass speakers (Fig 4*a*) used in high quality sound reinforcing systems to smaller horns (Fig 4*b*), handling only the speech range that may be used in noisy ball

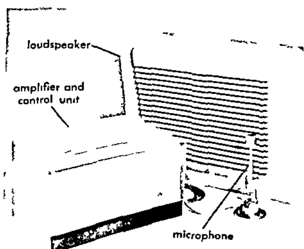


Fig 1 Simple form of public address system (Stromberg-Carlson Co, Division of General Dynamics Corp)

which enables him to achieve his maximal potential of usefulness to himself and society Prevention of disease and medical care made possible or implemented by organized social action are the substance of public health

The promotion of positive health is concerned with helping people achieve optimal well being and not merely absence of disease However almost nothing is known about the epidemiology of health Therefore public health programs for the promotion of positive health must await the fruits of epidemiological research

In broad terms then public health is concerned with determining the health status of the population (vital statistics) the genetic and bionomic factors in health and disease (epidemiology) and with building upon this knowledge the structure and function of organized health programs and services (public health administration) *See* BACTERIOLOGY MEDICAL EPIDEMIOLOGY HUMAN GENETICS MYCOLOGY MEDICAL VIRUS

[E M C]

Bibliography K F Maxcy *Rosenau s Preventive Medicine and Public Health* 8th ed., 1956 H S Mustard *An Introduction to Public Health* 3d ed., 1955

Puffin

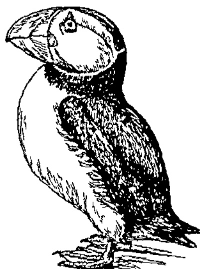
A bird *Fratercula arctica* a member of the family *Alcidae* along with the auks murres and auklets This short tailed short billed and rather chunky black and white diving bird breeds only in the extreme northern Atlantic and the adjacent Arctic Ocean area It walks well on its toes on land a trait not possessed by most of the family

Puffins nest in holes n he

fro

his

dog ey are also preyed upon by foxes



The Atlantic puffin *Fratercula arctica* length to 13½ in (From F L Palmer Fieldbook of Natural History McGraw-Hill 1949)

The puffin swims underwater with the use of its wings and easily catches fishes crustaceans and other food

The horned puffin *F corniculata* and tufted puffin *Lunda cirrhata* are birds of the North Pacific with similar habits *See* AUK CHARADRIIFORMES

[J D B]

Pulley

A wheel with a flat crowned or grooved rim used in conjunction with a rope belt or chain to transmit motion Pulleys for use with V belt and rope drives have grooved surfaces and are usually called sheaves A combination of ropes pulleys and pulley blocks arranged to gain a mechanical advantage as for hoisting is often referred to as block and tackle (*see* BELT DRIVE ROPE DRIVE)

Pulley construction Pulleys are made of cast iron steel wood or brass (Fig 1) Cast iron pulleys formed of one solid casting may have a solid or a split hub Those with solid hubs are held on the shaft w th ke s set s e o

six arms or spokes smaller ones may have only four arms those larger than 5 ft in diameter may have eight arms Split pulleys, formed of two sections bolted together at the hub and the rim are also used Either split pulleys or those with a split hub are preferred when the pulley face width is greater than 10 in When the pulley face width exceeds 20–24 in two sets of arms may be used to give better support of the rim Wood faced pulleys are lighter than cast iron and for flat belt drives have somewhat better power transmission characteristics Pulleys formed from sheet steel are light and free of the residual stresses that may be present in cast ones They are economical and have lower slippage than an equivalent cast iron pulley Brass may be used for small pulleys particularly for round belts and cords

The maximum safe rim speed for solid cast iron pulleys is in the order of 5000 feet per minute Split pulleys should be limited to 50–60% of this speed Built up steel pulleys while subject to some variation due to differences in design and construction should generally be limited to rim speeds of 6000 feet per minute

Pulley application When a belt drive must be capable of producing several different speeds of the driven shaft with a single speed of the driving shaft stepped or cone pulleys can be used (Fig 2) Diameters of such pulleys must produce the desired velocity ratio and maintain the necessary belt tension for any step Such drives are frequently made so that for a given driving pulley speed the speed of the driven pulley increases in a geometric ratio for each step When stepped pulleys are used in a crossed belt drive the sum of the diameters for any one step must be the same as for any other step In practice a pair of stepped pulleys is often designed to have the same dimen

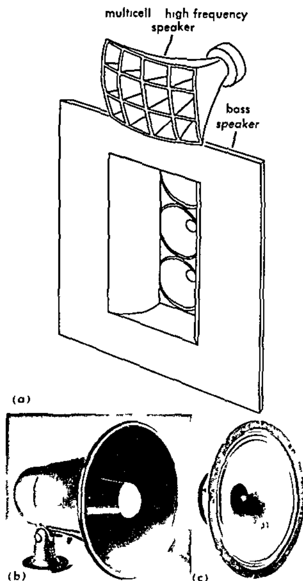


Fig 4 Loudspeakers used in PA systems (a) Multicell high frequency speaker with bass horn speaker for high quality sound reinforcing systems (b) Horn speaker used for the speech range in large and noisy locations (c) Small 8 in cone speaker used in distributed speaker systems (*Stromberg Carlson Co, Division of General Dynamics Corp*)

where the original sound source is in the same room as the loudspeakers When the amplification is set so that the sound from the loudspeakers arrives at the microphones at a pressure level equal to or higher than the sound pressure from the original sound source, the system breaks into oscillation The frequency of oscillation, or singing, will be close to any frequency or frequencies of exaggerated response in either the microphones or the loudspeakers Sound reinforcing systems that have exceptionally smooth frequency response characteristics are much less likely to have singing difficulties than those with irregular response Singing is easily prevented in a system by bringing the microphones as close to the original sound source and as far away from the loudspeakers as possible Directional loudspeakers placed so that they are pointed away from the microphones, and directional microphones with their directive beams pointed away from the loudspeakers, permit the relative distance from the microphone to original sound source to be increased

When widely spaced speakers are used, the finite velocity of sound and the successive time delays in the sound from the different speakers can cause severe garbling of speech when the delays are on the order of the time between syllables This type of intersyllable interference is often erroneously blamed on the speaker phasing Reversing the phase of the speakers does little to cope with the situation Rather the speakers must be oriented so the listeners cannot hear the sound from widely separated speakers In particularly critical sound systems a time delay is added to the electrical circuits feeding the speakers distant from the stage so the sound from the stage and that from the loudspeakers farthest from the stage arrive at the ears of the listeners together and not in sequence *See* SOUND REPRODUCTION SYSTEMS, ELECTRICAL

[F H SL]

Public health

Organized social action for the maintenance and promotion of health This broad concept has evolved from much more restricted definitions which limited public health to the actions of governmental health agencies directed at the prevention of infectious diseases The advances of medical science and the growing complexity of society have made more imperative the need for extensions of organized social action for the preservation of the individual's health

In the early days of public health, efforts were directed, often not too successfully, at controlling the epidemic spread of communicable diseases With the increase of bacteriological and epidemiological knowledge, the public health control of many communicable diseases has been placed on a firm scientific footing

The early emphasis of public health was on primary prevention of infectious disease by such measures as isolation of the sick and protective immunization, insect and rodent control water purification, and food and milk sanitation Today public health has grown to embrace activities directed toward primary prevention of noninfectious disease secondary prevention of both infectious and noninfectious disease, and care of the sick Possibilities of promoting positive health have also engaged the attention of public health workers

Epidemiological discoveries of the roles of environmental and cultural influences, of genetic determinants, and of somatic precursors have made it possible to prevent some of the noninfectious diseases Similarly it has been shown that early diagnosis and proper treatment of the incipient stages of certain diseases will prevent or minimize their progress The sick individual's need for medical care is obvious In such care there has been a growing emphasis on rehabilitation—the physical mental and psychosocial reeducation of the patient

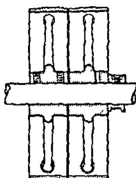


Fig 3 Tight and loose pulleys (J H Prageman Mechanism International Textbook 1943)

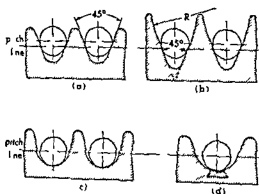


Fig 4 Pulleys for rope drives (a) Rope sheave (b) Multiple rope sheave (c) Idler sheave (d) Wire rope sheave (J H Prageman Mechanism International Textbook 1943)

rests at the bottom of an idler sheave (Fig 4c) A sheave for wire rope may have the bottom of the groove lined with rubber or a similar material (Fig 4d) [R C Y]

Pulmonata

A subclass of the gastropod mollusks containing the lung bearing land and fresh water snails Only a very few species live in salt water and these are intertidal These snails have lost their gills and in their place the mantle cavity has become a pulmonary sac which functions as a lung This organ consists of an invagination of the mantle which is lined with numerous blood vessels

Life cycle So far as known all pulmonates are hermaphroditic Mating usually occurs but self fertilization is also possible Land pulmonates have large eggs and lay 15 20 or as many as 400 at one time The eggs are usually buried in the soil or under leaves Fresh water pulmonates produce their eggs in a gelatinous mass which is attached to objects in the water such as leaves stones or even another egg mass All larval stages develop within the egg capsule and the young emerge as crawling snails

Habitat Land pulmonates occur in a great variety of habitats In tropical and subtropical areas many genera in several families live in trees feeding mainly on bark lichens Such tree snails usually become highly colored with characteristic color patterns for each genus Others are strictly calciphiles living only on limestone rocks and feeding on lichens Most of the land pulmonates remain on the ground usually hiding by day and feeding on decaying vegetation at night when the air is moist and cool In desert areas there are many types of land snail which have become adapted to living under very harsh conditions During long and protracted dry spells they bury themselves in the soil Having no operculum they produce an epiphragm a thin film of mucus which hardens leaving a small slit to admit the very little oxygen needed during a long period of inactivity The moment of rain in a desert area immediately brings them to the surface where they start feeding voraciously upon plant debris In less rigorous areas where there are seasonal wet and dry periods they may not enter the ground but cement themselves to trees and shrubs by means of the limy mucus and thus remain until they are again activated by rain

Terrestrial pulmonates Most of the land pulmonates possess shells a few such as *Limax* are without shells or the shell exists only as a small internal plate These shell less groups are usually referred to as slugs

Many land pulmonates are used as food for man particularly in Europe and North Africa The species eaten are mainly in the genera *Helix* and *Otala* and quantities are shipped to the United States and Canada particularly from North Africa to be sold mainly in the French and Italian sections of the larger cities

Fresh water pulmonates Fresh water pulmonates are found on all continents and most islands They are most abundant in regions of lakes and clear streams They are far less abundant in rivers and streams which carry any amount of silt Most of the fresh water pulmonates usually come to the surface periodically to breathe but many can remain below the surface obtaining their oxygen from plants or even directly from the water by the mantle the surface of which can act as a lung The genera *Physa Lymnaea* and *Helisoma* are among several of wide distribution See GASTROPODA [W J C]

Pulse generator

An electronic circuit capable of producing a waveform that rises abruptly maintains a relatively flat top for an extremely short interval and then rapidly falls to zero A relaxation oscillator such as a multivibrator (see MULTIVIBRATOR) may be adjusted to generate a rectangular waveform having an extremely short duration and as such it is a pulse generator However there is a class of circuits whose exclusive function is generating short duration rectangular waveforms These circuits are usually specifically identified as pulse generators An example of such a pulse generator is the trig

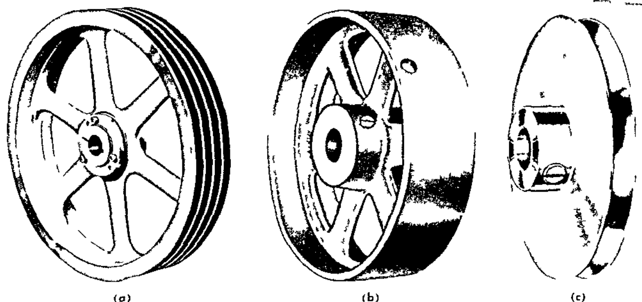


Fig 1 Typical pulleys (*a*) Multigroove V belt sheave (*Allis Chalmers*) (*b*) Solid hub crown face pulley (*Boston Gear Works*) (*c*) Brass pulley for round belt (*Boston Gear Works*)

sions When such pulleys (Fig 2) are used the relation of the driving shaft speed N and the speed of the driven pulleys on each side of the middle step is $N = (n_3 n_7)^{1/2}$

Speed variations of a belt drive can be obtained using speed cones rather than stepped pulleys These are similar to stepped cones except they are tapered smoothly rather than by steps The working diameter of the cone for any position of the belt is that at the middle of the belt The cones are laid out by calculating several diameters in the same manner as for step pulleys, plotting these equal distances apart along the pulley axis, and then fairing a smooth line through them When speed cones are used the belt must be guided as it approaches the cone to prevent it from climbing toward the large end of the pulley

With flat pulleys—those having the same diameter across the entire face—there is a tendency for the belt to run off the pulley if the shafts are slightly misaligned While guides at each edge of the belt as it enters the pulley can be used a crowned pulley will correct this problem with less belt wear A crowned pulley may be of either spherical or conical cross section (Fig 3) On such a pulley the belt tends to climb to the point of maximum diameter the center Thus it centers itself on the pulley The amount of crowning is usually small frequently being 1/96 of the pulley face width but values ranging from 1/40 for leather belts to 1/10 for cotton belts have been used Less crown is needed with higher belt speeds than with low speeds To prevent excessive travel of the belt back and forth across the pulley faces only one pulley in a pair should be crowned

Tight and loose pulleys (Fig 3) are used when a drive shaft is to remain in motion when the driven shaft is brought to rest The tight pulley is fastened to the driven shaft with a key or set screws The loose pulley, of a slightly smaller diameter, turns freely on the driven shaft but is kept in place by the hub of the tight pulley and a collar on the shaft The belt can be shifted while in motion, by a shipper that guides the advancing side of the belt onto the desired pulley

Sheaves for rope drives (Fig 4) such as those for hemp rope drives are similar to V belt pulleys, the rope wedges into the grooves and makes contact on the sides (Fig 4*a*) rather than at the bottom of the groove A curved groove is also used with a multiple rope system (Fig 4*b*) The rope

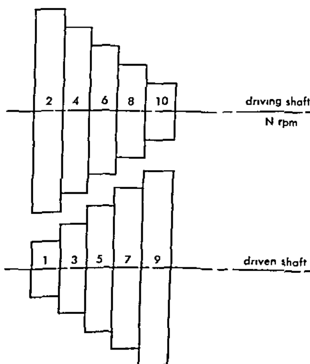
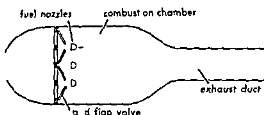
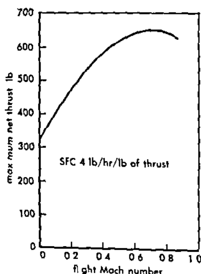


Fig 2 Equal stepped pulleys



Fg 1 Dagram of a pulse jet



Fg 2 Effect of flight Mach number on thrust of a typical pulse jet (German pulse jet length 137 2 in d ameter 21 6 in)

valves is the combustion chamber A fuel injection system is located at the entrance to the combustion chamber The chamber is also fitted with a spark plug Following the combustion chamber is a long exhaust duct which provides an inertial gas column

When the combustion chamber is filled with a mixture of fresh air and fuel a spark is discharged It ignites the fuel air mixture producing a pressure surge that advances upstream to slam shut the inlet valves and to block off the entrance Simultaneously a pressure pulse goes downstream to produce a surge of combustion products out the exhaust duct Thrust results from the rearward discharge of this gas at high velocity With the discharge of gas from the combustion chamber its pressure tends to drop Inertia causes the column of gas in the exhaust duct to continue to flow rearward even after the explosion pressure in the combustion chamber has been dissipated and this drops the combustion chamber pressure below atmospheric As a result the flap valves open and a fresh charge of air enters the combustion chamber As this air flows past the fuel nozzles it receives an injection of fuel and the mixture is then ignited by contact with the hot gas residue from the pre

Unlike the ramjet the pulse jet has an appreciable thrust at zero flight speed However, as the flight speed is increased the resistance to the flow of air imposed by the flap valves eventually causes substantial loss in performance and the pulse jet becomes less efficient than the ramjet

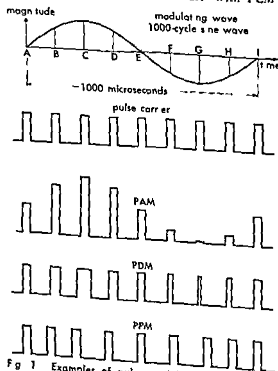
Failure of flap valves and valve seats by fatigue was found to be a problem Research has been conducted on valve systems other than that shown in Fig 1 and on valveless pulse jets

In addition to their use on the German V 1 buzz bomb pulse jets have been used to propel radio controlled target drones and experimental helicopters In the latter case they were mounted on the blade tips for directly driving the rotor The high fuel consumption noise and vibrations generated by the pulse jet limit its scope of application See PROPULSION [B PI]

Pulse modulation

A system of modulation in which the amplitude duration position or mere presence of discrete pulses may be so controlled as to represent the message to be communicated These several forms of pulse modulation are commonly called pulse amplitude modulation (PAM), pulse-duration modulation (PDM) pulse position modulation (PPM), and pulse code modulation (PCM), respectively For basic concepts technical terms and supplementary information see AMPLITUDE MODULATION, FREQUENCY MODULATION, MODULATION, PHASE MODULATION

Of all the different forms of pulse modulation PCM is the most outstanding This radically new form of pulse modulation represents a major contribution to the communications art With PCM



Fg 1 Examples of pulse modulat on (From H S Black Modulat on Theory Van Nostrand 1953)

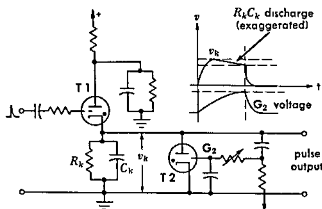


Fig 1 Thyratron pulse generator

gered blocking oscillator (*see* BLOCKING OSCILLATOR), which is a single vacuum tube or transistor relaxation oscillator having transformer coupled feedback from output to input

Pulse generators sometimes include but are usually distinguished from trigger circuits (*see* TRIGGER CIRCUIT) Trigger circuits by means of RC differentiating gated RLC peaking circuits or blocking oscillators generate a short duration fast rising waveform for initiating or triggering an event or a series of events in other circuits, such as monostable or bistable multivibrators In the pulse generator, the pulse duration and shape are of equal importance to the rise and fall times In this sense the blocking oscillator is a circuit which can be made to perform well in both respects

Thyratron pulse generator. An example of a pulse generator having a controllable pulsewidth is the thyratron pulse generator, shown in Fig 1 Tube T 1 fires when the positive triggering pulse is applied charging capacitor C_k to some positive potential If the resistances in the plate and cathode circuits are made sufficiently large, the current is not sufficient to maintain the discharge, and the potential v_k will slowly decrease as C_k discharges through R_k At the time the trigger is applied and for some time thereafter, the grid of T 2 is at a sufficiently negative value to prevent it from firing However, as the plate of T 2 rises after the trigger is applied, the grid rises more slowly, because of its RC time constants The grid ultimately becomes sufficiently positive to fire T 2, which then acts as a low impedance across C_k discharging it to near ground potential

Pulse forming networks. A network, formed in such a way as to simulate the delay characteristics of a lossless transmission line (*see* DELAY LINE), and appropriate switching elements to control the duration of a pulse form the basis for a variety of pulse generators

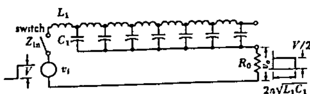


Fig 2 Principle of line-controlled pulse generator

A lossless transmission line has a characteristic impedance of

$$R_0 = \sqrt{\frac{L}{C}} \quad (1)$$

where L is the series inductance and C the shunt capacitance per unit length Such a line may be approximated by a network consisting of a number n of cascaded LC elements A pulse applied to the input of such a line reaching the output offers a time delay of

$$T_d - n\sqrt{LC} \quad (2)$$

An idealized circuit showing how such a network is used in a pulse generator is shown in Fig 2 A source is connected by a switch to a simulated unterminated transmission line in series with a resistance R_0 equal to the characteristic impedance of the line At the time voltage is applied and until all the capacitors in the line become fully charged the input impedance of the line is equal to R_0 and the current is equal to

$$I = \frac{V_s}{2R_0} \quad (3)$$

The step function progresses along the line charging each C in succession When it reaches the end it is reflected back with no change in phase and returns to the source in a time

$$T = 2n\sqrt{LC} \quad (4)$$

At this time the line is fully charged, the impedance becomes infinite, and current ceases to flow The pulse appearing across R_0 is suddenly terminated as shown For discussion of switching circuits suitable for supplying the line charging current *see* CLAMPING CIRCUIT, GATE CIRCUIT

Similar results can be obtained from the use of a current generator as a source and a short circuited line as the controlling circuit element

Various forms of delay line controlled pulse generators can be found, and some are capable of generating pulses containing considerable amounts of power for such applications as modulators in radar transmitters For an example of this type pulse generator, *see* BLOCKING OSCILLATOR, *see also* WAVE SHAPING CIRCUITS [C M C]

Bibliography B Chance et al (eds), *Waveforms*, 1949, G N Glasoe and J V Lebacqz, *Pulse Generators* 1948, L N Ridenour, *Radar System Engineering* 1947

Pulse jet

A type of engine widely known for its use during World War II on the German V 1 missile (Fig 1) The basic engine cycle was invented in 1908 The inlet end of the engine is provided with a grid to which are attached flap valves These valves are normally held by spring tension against the grid face and block the flow of air back out of the front of the engine They can be sucked inward by a negative differential pressure to allow air to flow into the engine Downstream from the fla-

Ease of multiplexing channels by time division is one of the important economic advantages of all forms of pul e mod ilation Circuits for accom pl hing this are simple and low in cost and because this low cost is shared by a numl er of chan nels the cost per channel is even lower When pulse modulation is applied to long distance communica tions these time division techniques also permit substantial simplification at so called way stations and branching points Because each information bearing pulse keeps its individuality in journeying from the first transmitter to the last receiver it is comparatively easy to drop and add message chan nels at various intermediate points along the way See TRANSMISSION THEORY AND METHODS

Basic concepts With the preceding description as background the fundamental aspects of pulse mod ilation which underlie so much of present day communication may now be presented in more de tail All forms of pulse modulation may be defined

ing of a series of regularly recurrent pulses In general the power associated with each pulse differs essentially from zero only during a limited interval of time which is the pulse width

Sampl ng Sampling is a process of extracting successive portions of predetermined duration taken at regular intervals from a continuously vary ing magnit ide time wave By sampling at a fast enough rate namely in excess of twice the highest significant frequency composing the sampled wave the samples will unambiguously define the wave And conversely given the samples the wave can be reconstr icted in all its detail

For example suppose the highest significant fre quency in a voice wave is less than 4000 cycles per second Then all the information necessary for its di tortionless reconstruction is given by short sam ples of the voice wave taken at regular intervals at the rate of 8000 samples per second that is by samples taken every 125 microseconds (μsec) This complete process including recovery of the voice wave i ill istrated step by step in Fig 2

A voice wave passes through a low pass filter (Fig 2a

The unit sampling function designated U is shown in Fig 2c This will be used presently to sample the voice Mathematically (Fig 2f) U is eq ial to a dc component k plus components at the sampl ng frequency f and its harmonics The inter val between pulses is $1/f$ and k is the ratio of pulse d iration t to the interval between pulses T

Next (Fig 2d) I is m iltiplied by I Because U is either unity or zero the product UI is a mathe matical process for sampling the voice The res ilt is a series of pos tive and negative pulses When U is unity the prod ct is I At all other times the prod ict is zero

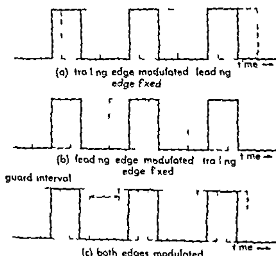


Fig 3 (a-c) Pulse-durat on modulation (PDM) (From H S Black Modulat on Theory Van Nostrand 1953)

Physically UI is an array of amplitude modu lated pulses Consequently an attenuated replica of I is obtained merely by passing UV through a low pass filter This may be demonstrated by per forming the indicated multiplication UI Amplifi cation restores the reconstructed wave to its origi nal value

A spectrum analysis of I and also UV is de picted by Fig 2g The top diagram is the spectrum of V The spectrum of UV is the spectrum of V small but exact plus upper and lower sidebands about f and about harmonics of f This illustrates in terms of the familiar concepts of amplitude mod ulation that passing UI through a low pass filter produces an attenuated replica of the sampled wave

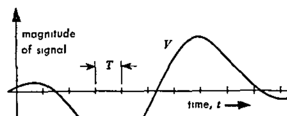
Pulse amplitude modulation (PAM) Pulse am plitude modulation is amplitude modulation of a pulse carrier The modulated wave (Fig 1) is lin early proportional to equally spaced samples of the modulating wave Another illustration is given in Fig 2d

Chief interest in PAM lies in its application to time division multiplexing Ordinarily the band width occupancy of PAM appreciably exceeds the theoretical minimum that is it appreciably ex ceeds the sum of the individual message bands Yet like other forms of ampl tude modulation PAM is not helped by wider bands unlike FM and unlike other forms of pulse modulation PAM cannot trade extra bandwidth for noise reduction

Pulse-duration modulation (PDM) Pulse dura tion modulation is modulation of a pulse carrier wherein the value of each instantaneous sample of a modulating wave produces a pulse of propor tional duration (Fig 3) by varying the leading trailing or both edges of the pulse PDM is also termed pulse length modulation or pulse width modulation. In Fig 3 solid lines den



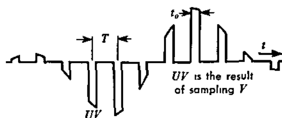
(a) source of voice input



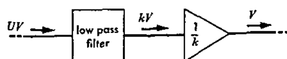
(b) typical voice input



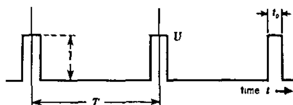
(c) diagram of U



(d) diagram of UV



(e) passing UV through a low pass filter and amplifier to obtain V



$$U = k + 2k \sum_{m=1}^{\infty} A_m \cos mCt$$

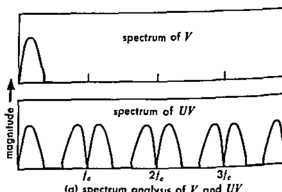
$$k = \frac{t_o}{T} = \text{duty cycle}$$

$$f_c = \frac{C}{2\pi}$$

$$\frac{1}{f_c} = T$$

$$A_m = \frac{\sin mk\pi}{mk\pi}$$

(f) enlarged diagram of U, unit sampling function



(g) spectrum analysis of V and UV

Fig 2 (a g) Properties of ordinary samples (*From H S Black, Modulation Theory Van Nostrand 1953*)

transmission circuits are simplified, over all trans mission losses are avoided crosstalk interference, and distortion are virtually eliminated, signals may be repeatered again and again without accumulating significant distortion, patterns of on or off pulses constitute the only type of signals that are propagated regardless of the type of message to be communicated, in the interest of reliability no on or off pulse can be lost (wrongly identified), and in the interests of efficiency and economy, little time is wasted

All forms of pulse modulation transmit message information intermittently rather than continuously Therefore unless the message information to be transmitted happens to be a time sequence of discrete values it must be reduced to this form before transmission Sampling (a process for obtaining a sequence of instantaneous values of a wave) accomplishes this by unambiguously representing a continuously varying wave by a series of distinct values (samples) Each message is momentarily sampled at regular intervals at a rate that is in excess of twice the highest message frequency to be communicated As will be explained presently in a PAM PDM or PPM system (Fig 1) a single pulse is used to specify the value of each sample

Fig 6 Pulse-code modulation (a) Quantizing of amplitude level and level designation by a code of pulses (b) Representation of a wave by a succession of coded pulse groups (From F E Terman Electronic and Radio Engineering McGraw Hill 4th ed 1955)

Clearly by using small enough steps errors may be cut to an arbitrary minimum (Fig 5)

When speech samples are quantized the steps are tapered as shown in Fig 5b By tapering most of the steps logarithmically nearly uniform percentage precision is obtained over most of the range and far fewer steps are required One hundred and twenty-eight logarithmic steps suffice to reproduce telephonic speech to a high degree of fidelity

Coding and multiplexing In PCM these 128 values are represented by a 7 pulse code Each voice sample is represented with adequate accuracy by some one of the possible patterns of seven on or-off pulses Channels are multiplexed by time division (Fig 6)

Similarly other forms of communication may also be represented by on or off pulses In practice a single PCM system may process many kinds of communications and may provide many channels of each kind

Ordinarily PCM systems are called upon to transmit large numbers of on-or-off pulses per second F [illegible]

each one way speech channel provided by PCM makes available a phenomenally fast one way data channel fast enough to operate simultaneously more than 1000 conventional teletypewriters In general the number of on-or-off pulses per second per message will depend upon the kind of message A one-way television signal might, for example, be represented by 70 000 000 on-or-off pulses per second

Repeatering By having to make only on-or off distinctions PCM is able to deliver a high quality signal even when noise and interference are so bad that it is barely possible to recognize the pulses At each regenerative repeater just so long as incoming pulses are correctly identified a new pulse generator may be caused to generate new pulses of correct magnitude waveform and timing

[H S BL.]

Bibliography L Marton (ed), *Advances in Electronics (and Electron Physics)* 2d ed vol 3 1957

Pulse modulator

A device for the pulse modulation of a radio-frequency carrier signal In pulse modulation information is transmitted by generating a train of discrete pulses whose amplitude duration, position or mere presence is controlled in accordance with the signal There are several forms of pulse modulation pulse amplitude modulation (PAM) pulse-duration modulation (PDM) pulse-position modulation (PPM) and pulse code modulation (PCM) In each case modulation is achieved by converting the continuous signal wave into a series of [illegible]

In the following discussion most of the circuits shown are idealized and simplified to illustrate the basic principles In a practical system such circuits are usually more complex

Pulse amplitude modulation (PAM) This requires least complexity of any of the pulse-modulation schemes and is usually used in the initial stages of other pulse modulation methods The amplitude of the input signal is sampled periodically at a fixed frequency producing a sequence of pulses whose amplitude varies in accordance with the modulating signal In a common practical

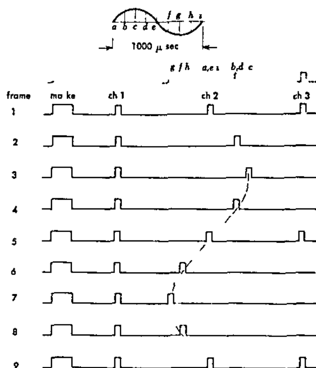


Fig 4 Pulse position modulation modulation of channel 2 by a sine wave (*From H S Black Modulation Theory Van Nostrand 1953*)

In contrast to PAM PDM *which was invented* by R A Heising in 1924 is able to trade extra bandwidth for noise reduction This noise advantage of PDM over PAM makes multiplexing by PDM even easier than multiplexing by PAM inasmuch as certain tolerances for controlling interchannel interference may be eased by an amount corresponding to the noise advantage However in order to achieve this important advantage the instantaneous values of interference must not be permitted to exceed the so called improvement threshold often enough to be disturbing

Only the position of the modulated edge or edges conveys information and the part of each PDM pulse that conveys no information represents wasted pulse power When this wasted power is subtracted from PDM the result is PPM which was invented by R D Kell in 1934 This power saving constitutes the theoretical advantage of PPM over PDM

Pulse position modulation (PPM) Pulse position modulation is modulation of a pulse carrier wherein the value of each instantaneous sample of a modulating wave varies the position in time of a pulse relative to its unmodulated time of occurrence

Figure 4 illustrates the sinusoidal modulation of the channel 2 pulse Successive diagrams indicate the change in the relative position in time of the channel 2 pulses from sample to sample for nine successive samples When a particular channel is idle that channel pulse recurs every 125 μsec When a channel is busy its pulse comes earlier or later depending upon the polarity of the sample The exact displacement of the pulse from its unmodulated position is proportional to the magnitude of the sample to be communicated All channel pulses are of constant magnitude and constant duration

Channel pulses one for each channel are transmitted in turn and are preceded by a synchronizing pulse called a marker This array of marker plus channel pulses repeats itself every 125 μsec and is called a frame In Fig 4 the synchronizing pulse is identified by its longer time duration Its function is to control the timing of the receiver with high accuracy

In practical applications even though PPM is more efficient than PDM both are highly inefficient when used for certain purposes for example when used for multiplexing ordinary telephone channels Consequently communication engineers have shown considerable interest in PCM which not only is more efficient but also possesses many other very important advantages

Pulse code modulation (PCM) Pulse code modulation (invented by H A Reeves in 1939) is a method of transmitting continuously varying message waves in which first the message wave is sampled second the value of each sample is replaced by the closest one of a finite set of permitted values and third these permitted values are then each unambiguously represented by some one of the possible patterns of N on or off pulses These three operations are known as sampling quantizing and coding respectively

Sampling Modern PCM is based upon the recognition that quantized samples may approximate an exact specification of a continuously varying wave as closely as desired Assume for purposes of illustration that a 7 pulse code is to be used This is a binary system whose capacity is 2^7 permitted or quantized values Each sample after being coded would be represented by some one of the 128 possible patterns of seven on or off pulses

Quantization Quantization changes a continuously varying signal to a stepped signal Graphically quantization means that the straight line representing the relationship between input and output has been replaced by a stair step function

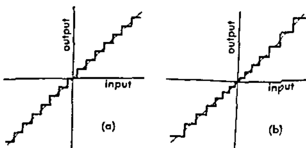


Fig 5 Relation between input and quantized output (a) Quantization uniform (b) Quantization tapered (*From H S Black Modulation Theory Van Nostrand 1953*)

Fig 2 Electronic distributor for a four channel time division multiplex transmitter (a) Transistor amplifier Q_1 together with switches Q_2 and Q_3, which permit input A to be connected to the output line The switches are operated by square waves of frequency f_s and $f_s/2$ (b) Waveforms needed to operate the switches

Pulse position modulation (PPM) PPM some-times is ...

f...

p... ...e of each pulse varies with the amplitude of the input signal Figure 4 shows one method of converting PDM to PPM together with representative waveforms The variable-width pulses are first differentiated and then rectified to permit only the negative peaks to trigger a monostable multivibrator In the latter each negative trigger pulse causes a pulse of constant amplitude and width to be generated whose time position is proportional to the amplitude of the modulating signal (input)

Pulse code modulation (PCM). PCM is a form of pulse comm...

sa...

is ...

a f...

ser... ...code pattern of ON or OFF pulses The three operations mentioned are referred to as sampling quantizing and coding, respectively For the advantages of this method and for further discussion *see* PULSE MODULATION

The sampling techniques needed for PCM can be similar to those employed in PAM Although electronic quantizing and coding often can become highly complex for the purpose of explanation a relatively simple scheme using eight linear quantizing steps and a beam coder tube will be described

The quantizing of the pulses generated by PAM can be explained with the aid of the diagrams

Fig 3 Conversion of PAM signals to PDM (a) Block diagram (b) Representative waveforms

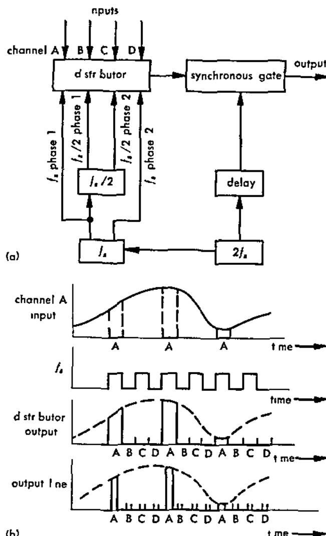


Fig 1 Time-division multiplex transmitter employing pulse-amplitude modulation (a) Block diagram showing the essential components in a four channel system (b) Waveforms associated with channel A

system the trains of pulses representing several input signals are interlaced to form a time-division multiplex system

Figure 1a is a block diagram of the essential parts of a four channel time division multiplex transmitter that uses PAM Figure 1b depicts some of the representative waveforms The principal elements of this system are the distributor or commutator which can be either electromechanical or electronic a source of square waves at the sampling frequency f_s and a source of square waves of one half the sampling frequency $f_s/2$ obtained from a bistable circuit arranged to provide both phases of the square wave The function of each is described below

Figure 2 shows a simplified schematic diagram of an electronic distributor which can be employed in a four channel time-division mutiplex PAM system For the purpose of explanation only the switching operation which occurs to connect the input channel A to the output line will be discussed When a positive voltage is applied to its emitter p n p transistor Q_1 becomes a normal amplifier and the signals appearing at its base are amplified and appear at the output However the positive voltage is applied to the emitter of Q_1 only during the periods when tandem switches Q_2 and Q_3 are both made conducting The sampling frequency square wave f_s of phase one alternately applies negative and positive voltages to the base of the transistor switch Q_3 causing it to open (conduct) and close respectively The $f_s/2$ square wave of phase one controls the transistor switch of Q_2 in a similar manner An examination of the waveforms in Fig 2b shows that there is only one interval during a distributor cycle when phases one of both square waves are negative At any other interval phase one of either one or the other of the two square wave frequencies is positive and the input A becomes isolated from the distributor output Inputs B C and D are connected to the output line in a similar manner Their associated switches are connected to the proper phases of the square-wave sources as indicated schematically in Fig 2a

In a practical PAM system it is necessary to provide guard time between pulses to reduce interchannel crosstalk Fig 1a shows the distributor output pulses passing through a synchronous gate operating at twice the sampling frequency This gate operating on the same principle as those used in the distributor proper selects the center portion of each pulse in the train The resultant output is a series of amplitude modulated pulses having only one half of their original width with blank time between each pulse The waveform in the output line with signals due to channel A alone is shown in Fig 1b

Pulse duration modulation (PDM) *PDM* can be produced by converting a train of pulses produced by a PAM system by means of a circuit such as that indicated in the block diagram in Fig 3 Figs 3a and 3b show that if the signals generated in a PAM system (waveform A) are added to constant amplitude sampling frequency pulses (waveform B) a train of pulses of varying amplitude but with a finite minimum value can be produced (waveform C) The resultant pulses can be differentiated and clipped resulting in a series of reverse saw tooth shapes (waveform D) in which the width of each saw tooth becomes proportional to the amplitude of the particular pulse When passed through a limiter the desired sequence of pulses having constant amplitude and varying in width in accordance with the signal is generated By properly adjusting the differentiating and clipping circuits the pulse corresponding to zero amplitude produces a pulse of minimum width which is used for reference

Guard time between pulses is required in a practical PDM system just as in the PAM Such circuits should be designed so that maximum amplitude samples when converted to pulse width do not occupy 100% of the available time slot In addition the differentiating circuit must be completely discharged during the guard time to prevent interaction between adjacent pulses

The circuit shown in the illustration indicates the basic push pull amplifier The interest in high fidelity audio amplifiers has lead to the develop ment of more complicated circuits In general

ground with capacitors and therefore the variation in screen voltage with signal voltage introduces degeneration which tends to make the operation

in both tubes are simultaneously zero This is cor rected by more advanced circuitry

Driver stages The input circuit shown in the illustration employs a transformer to produce the necessary phase inversion Because of the limited frequency response of the transformer and also the size of transformer required a transformer is rarely used Instead a vacuum tube phase inverter is employed Since some phase-inverter circuits have a gain considerably greater than unity the over all gain of the amplifier can be increased more than it could be if a transformer with a step up turns ratio were used *See* PHASE INVERTER

[H F K]

Pycnogonida

A subphylum of marine arthropods consisting of about 600 Recent and perhaps 1 Devonian species The Pycnogonida, or Pantopoda are commonly called sea spiders They are characterized by reduc tion of the body to a series of cylindrical trunk somites supporting the appendages a large special ized feeding apparatus called the proboscis gono-pores opening on the second joints of the legs and a reduced abdomen In many genera such as *Nym phon* (Fig 1) there are seven pairs of append ages of which the first four namely the chelifores palpi ovigers and first walking legs are on the first or cephalic segment. This segment also bears

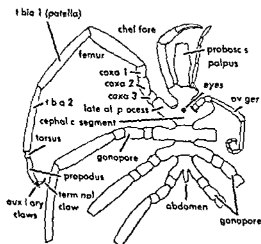


Fig 1 Diagram of a Nymphon show ng princ pal external features.

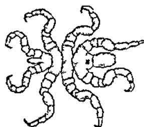


Fig 2 *Pycnogonum stearnsi* approximately 16 mm a common l ttoral pycnogon d of the Pac f c Coast

a dorsal tubercle containing *four simple eyes* Each of the remaining three trunk segments bears a single pair of legs In some genera the chelifores palpi or both are absent in the adult and the ovigers are lacking in the females Ovigers are al ways present in males and hold the ball like egg masses The extreme of reduction in which the adults lack anterior appendages except for ovigers in the male is represented by *Pycnogonum* (Fig 2), the common intertidal genus from which the group is named In some species there is an ad ditional trunk somite and pair of legs and one spe cies has 2 extra somites and 12 walking legs These polymerous forms occur in paucigeneric families with a large number of species to a genus and are not known in families with a large number of pau cispecific genera

Reduction of the body is accompanied by in creased length of the legs so that a typical pycnog onid appears to be a bundle of long legs These contain diverticula of the digestive tract and gonads and the sexual products ripen in the fem ora The proboscis is a rigid tripartite structure often longer than the combined trunk segments with a three cornered *mouth* at the apex, and con taining a straining apparatus which enables the animal to feed on the juices of soft bodied *animals* The nervous system follows the usual arthropod pattern of supraesophageal ganglion or brain circumesophageal ring paired ventral ganglia and ventral nerve cords The open circulatory system consists of a dorsal tubular heart with 2 3 pairs of lateral ostia Respiratory and excretory systems are lacking The femora of the males contain special ized cement glands which produce an adhesive sub-stance for forming the egg ball

Affinities The chelicerate condition of the first pair of appendages and the patellar joint of the legs are characters suggesting affinity with the Chelicerata but the multiple genital openings on the legs and the occurrence of polymerous forms are unique among arthropods (*see* CHELICERATA) The terminal joints of the oviger have intratarsal muscles another unique feature while the pre tarsus or propodus of the legs has both retractor and extensor muscles in common with both Cheli cerata and Crustacea The protonymphon larva (Fig 3) of the pycnogonida invites comparison with the crustacean nauplius larva but the three pairs of legs are not biramous and the re-

voir or, through long pipe lines, to a location at similar or different elevation In a power plant, pumps circulate cooling water or oil at low pressure and transfer water from heaters at moderate pressure to steam generators at pressures of several thousand pounds per square inch In chemical plants and refineries pumps transfer a great variety of fluids or charge them into reactors at higher pressure In hydraulic systems, pumps supply energy to a moving stream of oil or water, which is readily controlled, to move a piston or press platen or to rotate a shaft as required by the specific process *See* CENTRIFUGAL PUMP, DISPLACEMENT PUMP

Electromagnetic pumps. Where direct contact between the fluid and the pumping machinery is undesirable, as in atomic energy power plants for circulating liquid metals used as reactor coolants or as solvents for reactor fuels, electromagnetic pumps are used There are no moving parts in these pumps, no shaft seals are required The liquid metal passing through the pump becomes, in effect, the rotor circuit of an electric motor The two basic types are (1) conduction and (2) induction

The conduction type of pump which can be used with either direct or alternating current, confines the liquid metal in a narrow passage between the field magnets Electrodes on each side of this channel apply current through the liquid metal at right angles to the magnetic field and the direction of flow This current path is like the flow of current in the armature winding of a motor

The induction or traveling field type of pump operates only on polyphase alternating current The liquid metal is confined in a thin rectangular or annular passage thermally insulated from the slotted stator This stator with its windings is similar to the stator of a squirrel cage motor cut through on one side and rolled out flat The traveling field induces currents in the liquid metal similar to the currents in a motor armature *See* ELECTROMAGNETIC PUMPS [E F W]

Pumpkin

Two distinct definitions for pumpkin are recognized The first is restricted to varieties of the species *Cucurbita pepo* and *C moschata*, the second includes the edible fruit of any species of *Cucurbita* utilized when ripe as forage, as a table vegetable, or in pies All species belong to the plant order Campanulales

The second definition is more widely accepted Accordingly, the following popular varieties are classed as pumpkins *C pepo*, Connecticut Field Small Sugar, and Winter Luxury, *C mixta*, Cushaw, and *C moschata* Kentucky Field and Dickinson Canned pumpkin, however, is usually made from a blend of pumpkins and winter squashes Cultural practices are similar to those used for squash Harvesting generally begins when the fruits are mature usually four months after planting New Jersey, Illinois, and California are important producing states *See* CAMPANULALES, SQUASH, VEGETABLE GROWING [H J C]

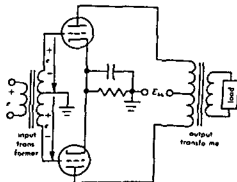
Push-pull amplifier

A two tube amplifier circuit often used as the power output stage of a multistage amplifier The power output stage in an audio amplifier is normally expected to furnish from 5 to 50 watts or more The use of one tube in this stage is not feasible, because the tube would have to operate Class A with a low plate circuit efficiency, on the order of 10% The use of two or more tubes in parallel does not improve the efficiency However, with two tubes operating in push pull, it is possible to supply the required amount of power with plate circuit efficiency on the order of 50% This higher efficiency means that the amplifier does not require as much power from the power supply, because less power is dissipated as heat by the power amplifier tubes *See* POWER AMPLIFIER

Operation. A simplified circuit for a push pull amplifier is shown in the illustration The input signals must be of equal magnitude but 180° out of phase The plate current in the tube with the positive signal is increasing while the plate current in the other tube is decreasing This relation between the plate currents gave rise to the name push pull Because of their phase relationship the two plate currents flowing in the two halves of the primary winding of the transformer produce an output similar to that of a single source connected to the transformer

A push pull amplifier may be operated Class A Class AB, or Class B (*see* AMPLIFIER) The greatest plate-circuit efficiency occurs when the operation is Class B Therefore, if large amounts of power are required, the amplifier is designed for Class B operation An additional merit of Class B operation is that the average current in the primary winding of the transformer is zero over one cycle of the input signal This feature makes it possible to design a smaller core for the transformer because there is no average flux level in the core *See* POWER AMPLIFIER

If the two tubes have identical characteristics and the transformer is considered ideal, the even harmonic components of distortion are absent in the output signal A high power Class B push pull amplifier can be designed with little harmonic distortion *See* DISTORTION (ELECTRONIC CIRCUITS)



Push pull amplifier

Jurassic and Cretaceous of the Northern Hemisphere They were formerly classified with other bilaterally symmetrical echinoids in the artificial assemblage Irregularia *See* DIADEMATACEA, ECHINOTHURIOIDA, IRREGULARIA [H B F]

Pyorrhea

An inflammation of the tissues surrounding the teeth principally the gums and dental periosteum It is marked by loosening of the teeth resorption of the surrounding bone and shrinking of the gums Pyorrhea accounts for most tooth loss in people over 35 years of age *See* TOOTH DISORDERS

Local causes include teeth irregularity, or malocclusion poor oral hygiene and mechanical irritation by dental plates or incorrect brushing Among the generalized disorders that may predis- ... are certain blood diseases faulty ... preg

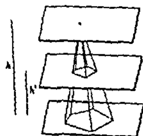
...

In pyorrhea a pocket forms ... and the gum border thus forming a receptacle for debris Progressive infection causes invasion of the periodontal tissue around the tooth with development of chronic inflammation Although bacteria are almost always present, they may not be fundamentally involved in originating the pyorrhea but will produce complications Early signs of pyorrhea are bleeding of the gums tooth mobility, engorgement of local blood vessels and gum changes Advanced or chronic pyorrhea is accompanied by severe bone resorption deep pus pockets around teeth and abscess formation

Early diagnosis and treatment of the condition as well as removal of the inciting cause if possible will generally produce a favorable outcome. Advanced cases, or those with extensive tissue alteration, may require extensive treatment [E G ST]

Pyramid and frustum

A pyramid is a polyhedron of which one face is called the base and the other faces (called lateral faces) are triangles having a common vertex which is called the vertex of the pyramid The distance from the vertex to the base is called the altitude The volume of a pyramid is one-third the product of its base times its altitude ($V = \frac{1}{3}Bh$) Sections of a pyramid formed by planes parallel to the base are similar to the base, and their areas are proportional to the squares of their distances from the vertex A pyramid with a triangular base is a triangular pyramid or tetrahedron If the base of a pyramid is a square (or any regular polygon) and the lateral edges are equal, the pyramid is called a square pyramid (or regular pyramid) The great pyramids of Egypt are square pyramids.



A regular pyramid

A frustum of a pyramid is a segment of a pyramid included between two parallel planes Its volume is given by the formula

$$V = \tfrac{1}{3}h'(B + \sqrt{Bb} + b)$$

where h' is the altitude of the frustum and where B and b are the areas of the two bases and also by the prismoid volume formula *See* POLYHEDRON, PRISMATOID AND PRISMOID [J S F]

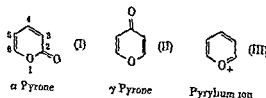
Pyramid of numbers

An expression of the idea that successive members of a given food chain within a community, become progressively fewer This concept was introduced by C. S Elton who showed that the relative decrease in numbers at each stage of the food chain is due to the facts that (1) smaller animals are preyed upon usually by larger animals and (2) smaller animals can increase faster than the larger and so are able to support the latter

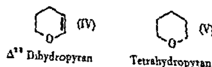
Because the number of individuals of a given species is not strictly related to that species' importance as a transformer of food the concept of the pyramid of numbers has been supplemented in quantitative ecology by those of the pyramids of biomass and of energy *See* BIOMASS, FOOD CHAIN [A MC]

Pyran

One of a group of organic compounds containing a heterocyclic six membered ring made up of one oxygen and five carbon atoms *See* HETEROCYCLIC COMPOUNDS Formulas (I) to (V) show pyrans of various kinds



α Pyrone γ Pyrone Pyrylium ion



Δ^{23} Dihydropyran Tetrahydropyran

All pyranosidic carbohydrates are tetrahydropyran derivatives. Glycals for example, glucal ... The α pyrone system ...

Recent families of Pycnogonida

Family	Chelifores	Palpi	Ovigers	
Nymphonidae	Present	5-jointed	10-jointed in both sexes	Includes the largest genus *Nymphon* with over 150 species and the polymerous *Pentanymphon*
Callipallenidae	Present	Lacking or reduced	10-jointed in both sexes	
Phoxichilidiidae	Present	Lacking	5–9-jointed in male only	
Endeidae	Absent	Absent	7-jointed in male only	Monogeneric paucispecific *Endeis*
Ammotheidae	Present, usually small and achelate	Well developed 4–10-jointed	9–10-jointed in both sexes	Many genera
Austrodecidae	Absent	Present 5–6-jointed	In both sexes reduced 4–7-jointed	Monogeneric *Austrodecus*
Colossendeidae	Lacking except in polymerous forms	Long 9–10-jointed	10-jointed in both sexes	Mostly deep-water forms with large proboscides. Genera are *Colossendeis* and polymerous *Pentacolossendeis*, *Decolopoda* and *Dodecolopoda*
Pycnogonidae	Lacking	Lacking	6–9-jointed in male only	Shallow-water to shore forms. Includes *Pycnogonum* and polymerous *Pentapycnon*

blances are superficial. The protonymphon has a number of specialized glands producing materials for attaching to or invading the host, and the structure of the proboscis is well advanced. The common name sea spider refers to the appearance of the adults and does not indicate affinity with terrestrial spiders.

Mode of life. Pycnogonids are found in all seas except the inner Baltic and Caspian, from intertidal regions to depths of 6500 m, and one species is bathypelagic at about 1000 m. They are especially common in polar seas. Most of the intertidal species spend their lives in association with some coelenterate as encysted parasitic larval and juvenile stages or are ectoparasitic as adults, being attached to anemones and hydroids by their claws and proboscides. A few have been found riding in hydromedusae and some occur in the mantle cavity of bivalves or on nudibranchs and holothurians. Most of the deep-sea species are known only as adults taken in deep-water dredge hauls, and their mode of life is a mystery.

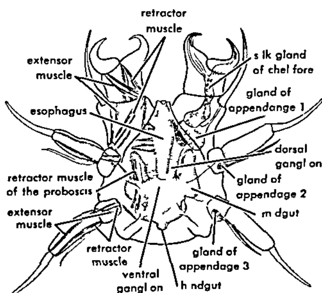


Fig. 3. Schematic diagram of Protonymphon larva showing muscles on left, digestive and nervous systems in center, and glands on right. (*Modified from H. Helfer and E. Schlottke*)

Classification. The Pycnogonida are classified primarily on the presence or absence of various anterior appendages; about 60 genera are recognized, grouped in 8 families. No ordinal distinctions can be recognized for Recent families, and many genera are somewhat artificially defined. In order to include the Devonian *Palaeopantopus* in the Pycnogonida, the Recent forms are assigned to an order Pantopoda and the fossil to the Palaeopantopoda. *See* ARTHROPODA; PALAEOISOPUS.

[J. W. H.]

Bibliography. J. W. Hedgpeth, On the phylogeny of the Pycnogonida, *Acta Zool. (Stockholm)*, 35:193–213, 1954; H. Helfer and E. Schlottke, *Pantopoda*, in H. G. Bronn (ed.), *Klassen und Ordnungen des Tierreichs*, vol. 5, pt. 4, 1935; O. W. Tiegs and S. M. Manton, The evolution of the Arthropods, *Biol. Rets.*, 33(3):255–337, 1958.

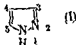
Pygasteroida

An order of Diadematacea which exhibits various stages in the backward migration of the anus out of the apical system. They have four genital pores (instead of five), noncrenulate tubercles, and simple ambulacral plates. All members are referred to a single family, the Pygasteridae. They apparently arose from Triassic Pedinidae and occur in the

many Guanajuato Mexico and Cobalt Ontario Canada *See* PROUSTITE SILVER METALLURGY
[C S HU]

Pyrazole

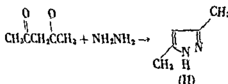
One of a group of organic heterocyclic compounds with two nitrogen atoms occupying adjacent positions in a doubly unsaturated five-membered ring A typical member of the group is pyrazole (I) *See*

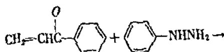


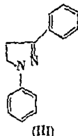
AZOLE HETEROCYCLIC COMPOUNDS The pyrazole system is resonance stabilized and aromatic in character Substitution occurs preferentially at the 4 position The nucleus is resistant to disruption by oxidation Certain drugs and dyes are pyrazole derivatives

The parent compound (I) is a water soluble colorless solid mp 70°C bp 187°C with an odor resembling that of pyridine Pyrazole is both a weak base (pK_a 2.53 at 25°) and a weak acid The hydrogen at position 1 can be replaced by potassium or by bromomagnesium

One general synthesis combines 1 3-dicarbonyl compounds with hydrazines for example the reaction of acetylacetone with hydrazine gives 3 5-dimethylpyrazole (II) A more common synthetic method condenses α β unsaturated aldehydes or ke

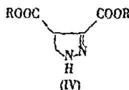




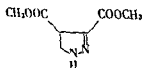


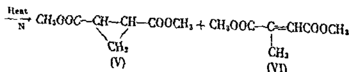
tones with hydrazines to give, not fully aromatic compounds but instead dihydropyrazoles or pyrazolines (III) Pyrazolines are also obtained when diazoalkanes react with olefinic compounds as in the

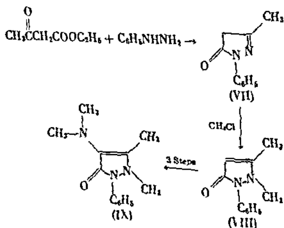


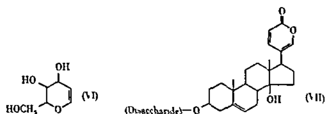


reaction of diazomethane with an ester of maleic acid to give a pyrazoline (IV) These pyrazolines can be dehydrogenated to pyrazoles Pyrazoles can be obtained directly when in the last two reactions acetylenic unsaturation replaces the olefinic unsaturation Pyrazolines are prepared by ring syntheses (III) and (IV) or by reduction of pyrazoles Pyrazolines are susceptible to oxidation. When there is no substituent on either nitrogen pyrolysis of pyrazolines gives rise to cyclopropanes or substituted olefins such as (V) and (VI)

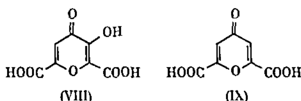






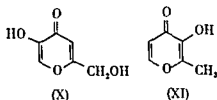


a variety of toad Meconic acid (VIII) from opium chelidonic acid (IX) from *Chelidonium majus*, and

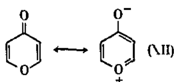


kojic acid (X) from *Aspergillus* molds are γ pyrones Maltol (XI) is a degradation product from carbohydrate materials

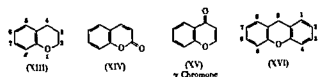
α Pyrones as lactones can be saponified and as *dienes can react with Diels Alder dienophiles Both* α and γ pyrones react with ammonia to give the corresponding pyridones γ Pyrones can be cleaved with alkali Other properties suggest that γ pyrones

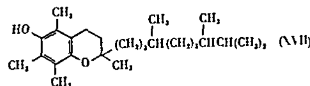


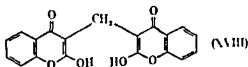
are resonance hybrids (XII)

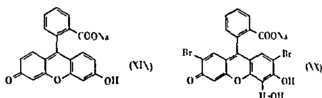


Formulas (XIII) to (XVI) show some benzopyran types *Vitamin* E, or α tocopherol (XVII) is a chroman (XIII) derivative Coumarin or α chromone (XIV), the fragrant body in clover, was formerly used in flavors and perfumes Dicoumarol (XVIII), originally encountered as the causative agent in hemorrhagic sweet clover disease of cattle now finds clinical use as a blood anticoagulant 2 Phenyl γ chromone derivatives are leaf pigments Fluorescein (XIX), used as a sea marker because of its intense greenish fluorescence eosin used as the dye in red ink, and mercurochrome (XX) used

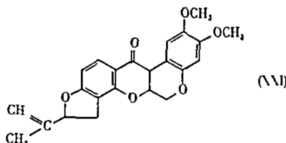




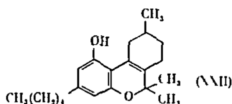




as a household antiseptic are all highly colored derivatives of xanthene (XVI) Rotenone (XXI)



an insecticide from *Derris elliptica* (Malaya Dutch East Indies) has three oxygen heterocyclic rings two of which are pyranoid The active ingredient in hashish as well as in marijuana—both from hemp (*Cannabis sativa*)—is tetrahydrocannabinol (XXII) a tricyclic pyran derivative



[W J CF]

Bibliography R C Elderfield *Heterocyclic Compounds* vol 1 1950

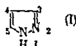
Pyrargyrite

A mineral having composition Ag_3SbS_3 and crystallizing in the hexagonal system Pyrargyrite crystals are prismatic with hemimorphic development and usually are distorted The mineral also occurs in massive form and in disseminated grains There is good *rhombohedral cleavage* the hardness is 2 5 (Mohs scale) and specific gravity is 5 85 The luster is adamantine and the color a deep ruby red to black giving it the name dark ruby silver Pyrargyrite is in places an important silver ore where it is found in veins associated with proustite and other silver minerals It has been mined as a silver ore at Chañarcillo Chile Freiberg Ge

many Guanajuato Mexico and Cobalt Ontario Canada *See* PROUSTITE SILVER METALLURGY [C S HU]

Pyrazole

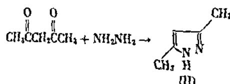
One of a group of organic heterocyclic compounds with two nitrogen atoms occupying adjacent positions in a doubly unsaturated five membered ring A typical member of the group is pyrazole (I) *See*



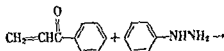
AZOLE HETEROCYCLIC COMPOUNDS The pyrazole system is resonance stabilized and aromatic in character Substitution occurs preferentially at the 4 position The nucleus is resistant to disruption by oxidation Certain drugs and dyes are pyrazole derivatives

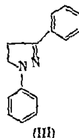
The parent compound (I) is a water soluble colorless solid mp 70°C, bp 187°C, with an odor resembling that of pyridine Pyrazole is both a weak base (pK_a 2 53 at 25°) and a weak acid The hydrogen at position 1 can be replaced by potassium or by bromomagnesium

One general synthesis combines 1 3-dicarbonyl compounds with hydrazines for example, the reaction of acetylacetone with hydrazine gives 3 5 di

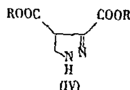


methylpyrazole (II) A more common synthetic method condenses α β unsaturated aldehydes or ke

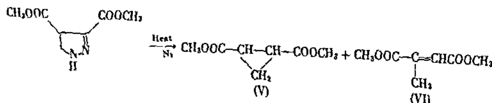


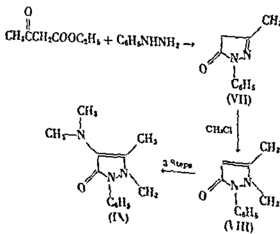


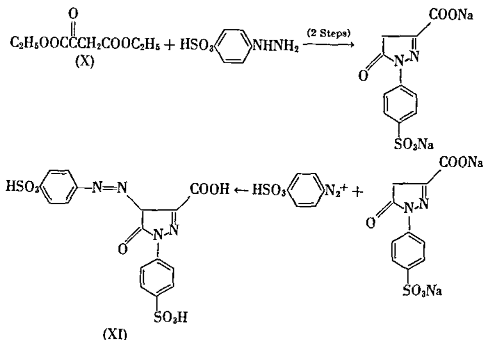
tones with hydrazines to give not fully aromatic compounds but instead dihydropyrazoles or pyrazolines (III) Pyrazolines are also obtained when diazoalkanes react with olefinic compounds as in the



reaction of diazomethane with an ester of maleic acid to give a pyrazoline (IV) These pyrazolines can be dehydrogenated to pyrazoles Pyrazoles can be obtained directly when in the last two reactions acetylenic unsaturation replaces the olefinic unsaturation Pyrazolines are prepared by ring syntheses (III) and (IV) or by reduction of pyrazoles Pyrazolines are susceptible to oxidation When there is no substituent on either nitrogen pyrolysis of pyrazolines gives rise to cyclopropanes or substituted olefins such as (V) and (VI)







Antipyrine (VIII) and aminopyrine (IX) are pyrazole derivatives of considerable value as an analgesic and an antipyretic, respectively The large scale synthesis of these two materials starts with the condensation of phenylhydrazine with acetoacetic ester to give 1 phenyl 3 methyl 5 pyrazolone (VII), also known as methylphenylpyrazolone or as Developer Z. Methylation of (VII) gives antipyrine (VIII) Subsequent nitrosation, reduction (with zinc), and dimethylation leads to aminopyrine (IX)

Azo coupling of pyrazolone (VII) or of related compounds gives 4 azopyrazolone derivatives, which are of interest as wool, food and photographic dyes The synthesis of one such dye (XI) starts with oxaloacetic ester (X) and proceeds as indicated [W J CE]

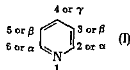
Bibliography R C Elderfield, *Heterocyclic Compounds*, vol 5 1957

Pyrethrum

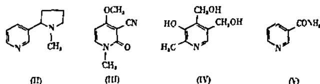
Flowers of the genus *Chrysanthemum*, used in making insect powder Persian insect flowers, *Chrysanthemum coccineum*, Caucasian insect flowers *C marschallii* and the Dalmatian insect flowers, *C cinerariaefolium*, are the species of greatest importance The Dalmatian plant, native to Dalmatia in Yugoslavia has long been cultivated and is the chief source of commercial insect flowers The partially opened flower heads are used in the preparation of a powder, which is the potent insecticide *See* CAMPANULALES, INSECTICIDE [P D S]

Pyridine

An organic heterocyclic compound containing a triunsaturated six membered ring of five carbon atoms and one nitrogen atom *See* HETEROCYCLIC COMPOUNDS Pyridine (I) and pyridine homologs are obtained by extraction of coal tar or by synthesis The following are available in commercial quantities pyridine 2, 3-, and 4-methylpyridine (also known respectively as α, β, and γ picoline) 2,4-dimethyl, 2 6 dimethyl, and 3 5 dimethylpyridine (also known respectively as 2 4-, 2,6-, and 3 5



lutidine) 2 methyl 5 ethylpyridine (also called aldehydecollidine), and 2,4 6 trimethylpyridine (also called 2 4 6 collidine) Other pyridine derivatives produced on a large scale include nicotinic acid (pyridine 3 carboxylic acid) for preparation of nicotinamide nicotine (II) for its insecticidal

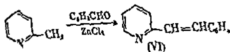


properties, 2 aminopyridine for synthesis of medicinals piperidine (hexahydropyridine) as a solvent, and 2 vinylpyridine as a polymerizable monomer The pyridine system is found in natural products, for example, in nicotine (II) from tobacco in ricinine (III) from castor bean in pyridoxine (IV), in nicotinamide or niacinamide (V), and in several groups of alkaloids

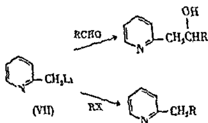
Properties Pyridine (I) is a colorless hygroscopic liquid with a pungent unpleasant odor It boils at 115 2–115 3°C, its density (20/4) is 0 98272 and n_D^{20} is 1 50920 Pyridine is miscible with organic solvents as well as with water A constant boiling mixture, bp 92°C, forms with three molecules of water Dry pyridine is obtained by treatment with barium oxide followed by calcium hydride or phosphorus pentoxide Pyridine is a tertiary amine (pK_a 5 17 at 25°) that combines ....

y with Bronsted and Lewis acids The pyridine stem is aromatic It is stable to *heat to acid* nd *to alkali* It undergoes substitution, with the position favored in sulfonation and nitration It hows *resonance* energies of 21 31 kcal/mole Pyri ... used as a solvent for organic and inorganic

tion at the methyl group of 2 and 4 ... tends to occur more readily than at the methyl



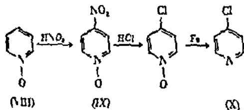
not react With butyllithium or sodium amide 2 and 4-methylpyridines metalate to give pyridyl methyl metals (VII) which react normally with



carbonyl compounds and with alkyl halides

Many halogenated derivatives of pyridine are known Direct bromination at 300–400°C places

... aminopyri d ne or by mercurating and then brominating *pyri d ne* 2 Chloropyridine is formed when *N* methyl 2 pyridone is treated with phosphorus pentachlo ride 4-Chloropyridine (X) is prepared by nitra



ting pyrid ne N-oxide (VIII) exposing the 4-n tropyrid ne-N-oxide (IX) to the action of concen trated hydrochloric *acid* and removing the oxide oxygen by iron acetic acid reduction The 2 and 4-halo substit ients are more readily replaced by hydroxy alkoxy and amino than the 3-halo sub-

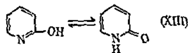
stituent Bromopyridines form Grignard and lith ium derivatives that react normally

3-Nitropyridine from nitration of pyridine can be con... dine-N ... 4-amin... industrial scale by direct amination of pyridine ... the man ifacture of the ... and the bac

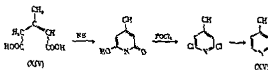


In most of its reactions 3-hydroxypyridine be haves as a normal phenol just as 3-aminopyri dine behaves as a normal aromatic amine How ever hydroxy or amino groups at the pyridine 2 or 4 positions show some reactions that are not char acteristic of phenols or aromatic amines Hydroxyl or amino groups on any pyridine position make *electrophilic* substitution easier and as ortho-para directing groups take control of the orientation 2-Pyridone is prepared from 2 aminopyridine by a diazot ...

... 2 6-dihydroxypyridines The last two pyridine derivatives exist almost en tirely in their tautomeric forms, that *is* as 2 pyri done (*XIII*) and 6-hydroxy 2 pyri*done* respec

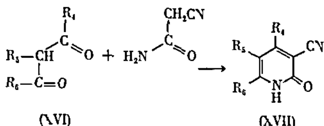


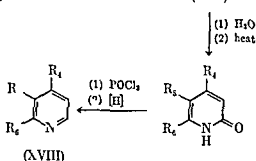
tively Glutaconic acids cyclize with ammonia to give 6-hydroxy 2 pyridones The oxygen at the 2 and 6 positions can be removed by standard con versions with *phosphorus* oxychloride to the 2 6-dichloro derivative followed by reductive dechlori nation *In this* way for example β methylgluta *conic* acid (*XIV*) can be converted to 4-methylpyri



dine (XV) 2 Pyridones (*XVII*) *are formed* when 1 3-*dicarbonyl* compounds (*XVI*) *or* their equiva lents react with cyanoacetamide *Subsequent steps remove* the cyano groups *as well* as the oxygen from (XVII) to furnish substituted pyridines (*XVIII*)

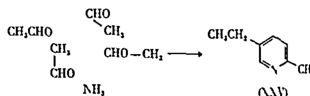
1,5 D carbonyl compounds or their equivalents cyclize with ammonia to give pyridines *Thus the*



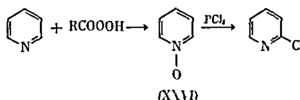


diketone (XIX) from acetoacetic ester and ethyl orthoformate gives the pyridine (XX) In a closely related process the reaction of ethyl β aminocrotonate (XXI) with ethoxymethyleneacetoacetic ester (XXII) gives the same product The Hantzsch synthesis combines four molecules to form an intermediate dihydropyridine derivative (XXIII) which can be readily oxidized to the corresponding completely aromatic derivative (XXIV) The Chichibabin aldehyde ammonia synthesis involves the condensation of ammonia with aldehydes and ketones Generally mixtures of pyridines are obtained with the course of the reaction depending on such factors as nature and proportion of reactants reaction time temperature and catalyst In a commercial process 2 methyl 5 ethylpyridine (XXV) is obtained in unusually high yields (60–70%) from acetaldehyde and ammonia Presumably synthetic industrial pyridine and its homologs are prepared by the Chichibabin or some related process

Derivatives Pyridine compounds containing positively charged nitrogen include simple and quaternary pyridinium salts acylpyridinium salts

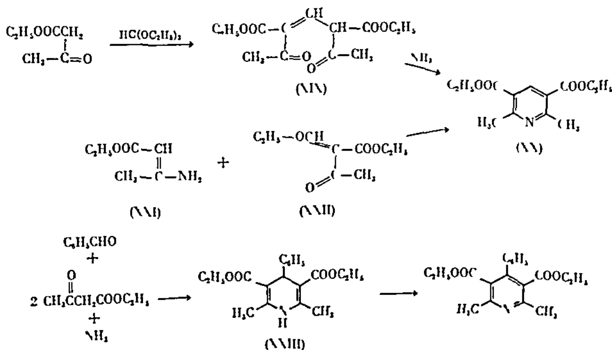


and pyridine *N* oxides The reaction of pyridine with alkylating agents gives quaternary salts crystalline solids whose aqueous solutions conduct electricity The quaternary salts are readily oxidized by alkaline ferricyanide to the *N* substituted 2 pyridones With acyl halides pyridine form *N* acylpyridinium salts which are powerful acylating agents for OH NH and SH groupings Pyridine-*N* oxide (XXVI) prepared by oxidation of pyridine with

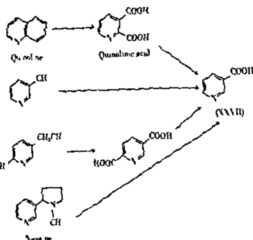


organic peracids reacts with phosphorus pentachloride to give 2 chloropyridine and with acetic anhydride to give 2 acetoxypyridine Pyridine-*N* oxides nitrate at the 4 position and by so doing provide a route to 4 substituted derivatives

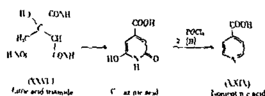
Pyridine aldehydes are prepared by oxidation of groups already on the ring Acetylpyridines can be synthesized from pyridine carboxylic esters and ethyl acetate by the Claisen condensation The reactions of pyridine aldehydes and ketones are normal



Many pyridine carboxylic acids are known Their reactions as dipolar materials are not exceptional Thermal decarboxylation is a standard process with loss of carboxyl from position 2 easier than from 3 or 4 and loss of carboxyl from position 4 easier than from 3 Nicotinic acid (XXVII) is man

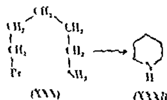


ufactured by oxidation of 3-methyl pyridine (β picoline) or nicotine it is also obtained by *oxidation of either 2 methyl 5-ethylpyridine or quinoline followed* by decarboxylation of the resulting pyridine dicarboxylic acids Pyridine 2 carboxylic *acid* (*a picolinic* acid) can be prepared by oxidation of 2 methylpyridine or by *carbonation of* 2 pyridyllithium Pyridine 4-carboxylic acid (isonicotinic acid) is obtained by *oxidation of* 4-methylpyridine or by synthesis from citric acid See (XXVIII) *to* (XXIX) The acid hydrazide of



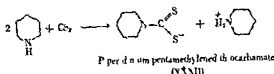
isonicotinic acid (Isoniazid) is a tuberculostatic agent

Although dihydro- and tetrahydropyridines are known the hexahydropyridines or piperidines are the most common reduced forms of pyridine The piperidines may be prepared by reduction of pyridines or by cyclization of bifunctional compounds for example the conversion of 5 bromo-1 aminopentane (XXX) to piperidine

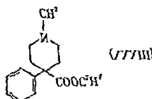


Piperidine (XXXI) the parent *compound* is a colorless unpleasant smelling liquid (b p 105 6°C) completely miscible with water The general properties of piperidine are those of a normal secondary

gated in connection with naturally occurring amino acids (for example pipecolinic acid is piperidine 2 carboxylic acid) as well as with the degradative and synthetic chemistry of quinine The reaction product (XXXII) from piperidine and carbon di



sulfide is a rubber accelerator 4 4-Disubstituted piperidines such as Demerol (XXXIII) *are* anal



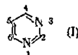
gesics

See SULFA DRUGS [W.J CE]

Bibliography R C Elderfield (ed) *Heterocyclic Compounds* vol 1 1950 E H Rodd (ed) *Chemistry of Carbon Compounds* vol 4A 1957

Pyrimidine

An organic heterocyclic compound in which a six membered ring contains nitrogen atoms at the 1 and 3 positions (I) *See* DIAZINES HETEROCYCLIC COMPOUNDS Pyrimidine compounds include vita



min B_1 *as well as several of the heterocyclic bases* in nucleoproteins Pyrimidines accordingly are involved in fundamental biochemical processes Other important pyrimidines are found in synthetic medicinals such as ...

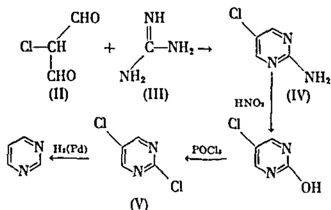
Properties

basic, water

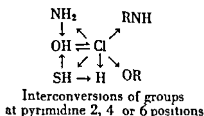
mp 21 7°C.

compound has been synthesized by ... ie parent

... yuloxy 5 chloro pyrimidine with phosphorus oxychloride gives 2 5-dichloropyrimidine (V) which furnishes pyrimidine on catalytic hydrogenolysis Most pyrimidine syntheses follow this same general pattern A 1 3-*dicarbonyl* compound (such as malonic ester a β keto ester β diketone β dialdehyde or their



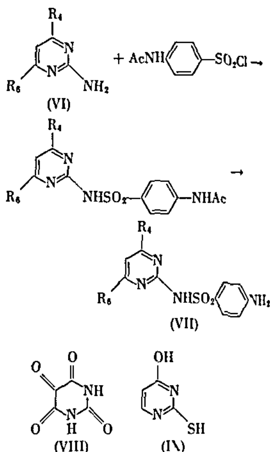
equivalent) or a β cyano carbonyl compound is condensed with guanidine or a related compound, such as urea, thiourea and amidines Accordingly, most pyrimidine ring syntheses furnish products with one or more hydroxyl, amino or sulfhydryl groups at positions 2 4 or 6 The accompanying diagram shows how these groups may be manipu lated in useful conversions



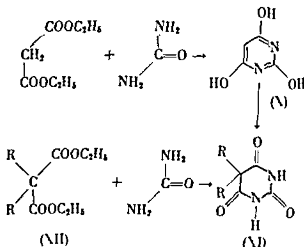
Interconversions of groups at pyrimidine 2, 4 or 6 positions

Pyrimidines are more stable to acids than to strong alkali The compounds resist disruption by oxidation They can be reduced to hydropyrimi dines Pyrimidine is a resonance stabilized mole cule structurally related to benzene However, car bon atoms 2 4 and 6 are adjacent to nitrogen so *that they are more susceptible to attack by nucleo*-philic agents than by electrophilic agents Thus 4-methylpyrimidine with sodium amide gives 2 amino 4-methylpyrimidine Other related conse quences of this disposition of the heteronitrogens are found in the facile interconversions of groups at the 2 4 and 6 positions in the enhanced reac tivity of methyl groups at these positions and in the relative ease of decarboxylation of pyrimidine-2 4 or 6 carboxylic acids Actually, only position 5 reacts with the familiar benzene substituting (electrophilic) reagents

Important derivatives. Three synthetic pyrimi dines (VII), sulfadiazine ($R_4 = R_6 = H$), sulfa merazine ($R_4 = CH_3$ $R_6 = H$), and sulfamethazine ($R_4 = R_6 = CH_3$), are important members of the sulfa family These compounds are prepared from the appropriate 2 aminopyrimidine (VI) Certain 2 4-diaminopyrimidines have antifolic acid activity 5 (p Chlorophenyl) 2 4-diamino-6-ethylpyrimidine has shown promise as an antimalarial drug Alloxan (VIII), administered either orally or intravenously, produces diabetic symptoms The effect is counter acted with 2 thiouracil (IX) 6-Propylthiouracil is a clinically useful antithyroid agent *See* FOLIC ACID, SULFA DRUGS

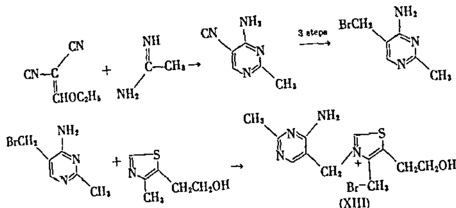


The barbiturates the 5 5 disubstituted barbituric acids (XI), acting chiefly as central nervous sys tem depressants are hypnotics and sedatives They are prepared by dialkylation of barbituric acid (X), which is obtained from malonic ester and



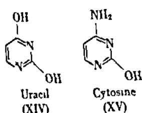
urea or by condensation of a disubstituted malonic ester (XII) with urea Some familiar and useful barbiturates (XI) are barbital or Veronal (5 5 diethyl), phenobarbital or Luminol (5 ethyl 5-phenyl), amytal (5 ethyl 5 isoamyl) and Nembutal (5 ethyl 5 α methylbutyl)

Thiamine or vitamin B_1 (XIII) the specific agent against beriberi is a pyrimidine One preparation starts by condensation of acetamidine and ethoxy methylenemalononitrile and proceeds as shown to give the synthetic vitamin (XIII)

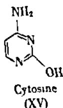


(XIII)

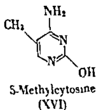
Nucleic acids are polymers the monomers of which are phosphorylated sugars (specifically ri bose or 2-d a heteron to (XX)



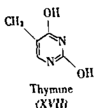
Uracil
(XIV)



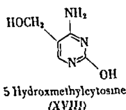
Cytosine
(XV)



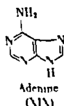
5-Methylcytosine
(XVI)



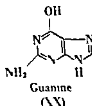
Thymine
(XVII)



5 Hydroxmethylcytosine
(XVIII)



Adenine
(XIX)

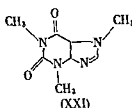


Guanine
(XX)

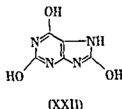
called nucle ides The phosphorylated sugar pyri mid ne combinat ons are called nucleotides Ac cord ngly nucleic acids are polynucleotides Nu cleo pr tein composed of nucleic acid plus protein is the building material of chromosomes and as such is of paramount importance in mitosis and in genetics Viruses are nucleoprotein in nature See NUCLEOPROTEIN

Purines of which adenine (XIX) and guanine (XX) are examples contain fused pyrimidine and *imidazole rings The* purines are important, not only as constituents of nucleic acid but also as the

phosphorylating coenzyme adenosine triphosphate (ATP) and the acetyl transferring coenzyme co-enzyme A Caffeine (XXI) a central nervous system stimulant is a substituted purine Uric acid or 2 6 8-trihydroxypurine (XXII) is obtained from



(XXI)



(XXII)

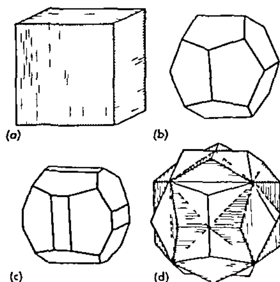
guano the excrement of birds and reptiles Deposi *tion of uric acid in the joints of man gives* rise to the abnormal condition known as gout See PYRI DINE [W J CE]

Bibliography R C. Elderfield (ed) *Heterocyclic Compounds* vol 6 1957

Pyrite

A mineral having composition FeS_2 and crystallizing in the isometric system Pyrite or iron pyrites is more commonly well-crystallized than any other sulfide mineral the cube is the dominant form Striae are usually present on the cube faces running at right angles to the edges The pyritohedron and octahedron are frequently present A penetration twin of two pyritohedrons is known as the iron cross Pyrite is also massive granular and stalactitic

Pyrite has a hardness of 6–6 5 (Mohs scale) and a specific gravity of 5 02. The luster is metallic and the color brass yellow in very fine grained com



Pyrite crystals (a) Cube (b) Pyritohedron (c) Combination of cube and pyritohedron (d) Penetration twin or iron cross (From C S Hurlbut Jr Dana's Manual of Mineralogy 16th ed Wiley 1952)

pact aggregates the color may be greenish Pyrite is the most common fool's gold but is hard and brittle whereas gold is soft and sectile Its high hardness also distinguishes it from softer chalcopyrite In the pure mineral iron makes up 46 6% and sulfur 53 4% When ignited the high percentage of *sulfur permits pyrite to support* its own combustion Small amounts of nickel and cobalt may be present and some analyses show considerable nickel indicating the possibility of a complete solid solution series between pyrite and bravoite $(Ni,Fe)S_2$ Gold copper nickel and arsenic reported in some analyses are probably the result of mechanical mixtures of other minerals and the elements present are not substituting for iron or sulfur in the structure FeS_2 is dimorphous and crystallizes in an orthorhombic modification as marcasite Marcasite is *distinguished from pyrite by its* lighter color lower specific gravity and crystal form *See* MARCASITE

Pyrite is the most common as well as the most widespread of the sulfide minerals It occurs under almost all conditions of mineral deposition from the high temperatures of an igneous magma to the temperatures of the ocean bottom near 0°C It is present as an accessory mineral in many igneous rocks and in some has formed large deposits as a magmatic segregation Large masses are found in *contact metamorphic* ore deposits and are mined for the associated copper minerals chiefly chalcopyrite and chalcocite Pyrite is a common mineral in most sulfide veins whether formed at great depths and *high temperature or shallow depths and low* temperature In veins it is associated with many minerals but most frequently with chalcopyrite sphalerite and galena Pyrite of both primary and secondary origin is common in sedimentary rocks and forms nodular and banded masses in coal known as brasses It is also found in metamorphosed sedimentary rocks in places in well formed crystals. *See* CHALCOPYRITE GALENA SPHALERITE

Pyrite is so universal in its occurrence that only a few major localities will be noted here Fine crystals have been found at Waldenstein Austria St Gotthard Switzerland on the island of Elba at Saint Pierre de Mesage France and in Cornwall England Large deposits occur at Rio Tinto and elsewhere in Spain where the pyrite and associated chalcopyrite are mined on a large scale In the United States fine crystals have been reported from many localities notably Rossie New York Leadville Colorado Clifton Arizona and Bingham Canyon Utah Well formed cubes are found in a chlorite schist at Chester Vermont Large massive deposits are found at Louisa Virginia Ducktown Tennessee, and Rico Colorado

Under oxidizing conditions pyrite readily alters *to various iron sulfates and eventually to goethite* and limonite These iron oxides form the chief minerals of the gossan which is the surface expression of pyrite-rich mineral deposits *See* ORE AND MINERAL DEPOSITS

Because of the high percentage of sulfur (53 4%) in pyrite this mineral is one of the sources of sulfur used in the manufacture of sulfuric acid In places it is mined for sulfur alone elsewhere sulfur is recovered as a by product in smelting of ores rich in pyrite In the United States native sulfur supplies most of the demand for the element Elsewhere however pyrite assumes a more important role as a source of sulfur *See* GOETHITE, LIMONITE SULFUR [C S HU]

Pyroclastic rocks

Rocks of extrusive (volcanic) origin composed of rock fragments produced directly by explosive eruptions Pyroclastic fragments may represent shattered and comminuted older rocks (volcanic plutonic sedimentary or metamorphic) or solidified lava droplets formed by violent explosion *See* TUFF *see also* VOLCANO [C A CA]

Classification of pyroclastic materials

Grain size mm	Unconsolidated rock	Consolidated rock
< 4	Ash	Tuff
> 4 < 32	Lapilli	Lapilli tuff
> 32	Blocks (angular)	Breccia
	Bombs (rounded)*	Agglomerate†

* Masses erupted as large blobs of liquid lava
† Essentially a breccia with abundant bombs and huge blocks, found within or near the volcanic vent

Pyroelectricity

A state of electric polarity *produced in certain* crystals by change of temperature Certain dielectric (electrically nonconducting) crystals develop an electric polarization (dipole moment per cubic *centimeter) ΔP when they are subjected to* a (uniform) temperature change ΔT For a small change ΔT the components ΔP of the polarization vector are given by

$$\Delta P_i = p \,\Delta T \qquad i = x\ y\ z$$

This effect is called the pyroelectric effect and the coefficients p the pyroelectric coefficients The necessary and sufficient condition for the effect is ... of a unique polar axis It is fulfilled

l thi im sodi im sulfate ...

The magnitude of the pyroelectric effect depends upon whether the thermal expansion of the crystal is prevented by clamping or whether the crystal is mechanically unconstrained In the clamped crystal one observes the primary pyroelectric effect whereas in the free crystal a secondary pyroelectric effect is superposed upon the primary effect The secondary effect may be regarded as the piezo electric polarization arising from thermal expansion (see PIEZOELECTRICITY) Hence the coefficients p for the secondary pyroelectric effect have the same temperature dependence as the coefficient ... x es ire proportional to T^4 (T in degrees absolute) whereas the coefficients of the primary effect appear to be proportional to T^2

From the standpoint of molecular theory the primary effect arises from a change of the electron distribution in the crystal whereas the secondary effect is mainly the result of the relative displacements of positive and negative ions Pyroelectric crystals can be regarded as having a built in or permanent electric polarization When the crystal is held at constant temperature this polarization does not manifest itself because it is compensated by free charge carriers that have reached the surface of the crystal by conduction through the crystal and from outside However when the temperature of the crystal is raised or lowered the permanent polarization changes and this change manifests itself as pyroelectricity There is no way to determine the magnitude of the total permanent polarization except for those special pyroelectric crystals called ferroelectrics in which the polarization can be reversed by an electric field See FERROELECTRICS

In a typical pyroelectric crystal such as tourmaline a temperature change of 1C° produces at room temperature a polarization of about 10^{-9} coul ... m

It is a consequence of thermodynamics that the pyroelectric effect has an inverse the linear electrocaloric effect A temperature change ΔT results when the permanent polarization is altered by an externally applied electric field ΔE

$$\Delta T = \frac{T}{\rho c_p} \sum p_i E_i \qquad i = x\ y\ z$$

In this expression ρ is the density c_p the specific heat at constant pressure p_i are the pyroelectric coefficients and T is the absolute temperature

The temperature changes that can be realized in typical pyroelectrics are of the order of magnitude 0 01°C [W K]

Bibliography M Born On the quantum theory of pyroelectricity *Revs Mod Phys* 17(2 3) 245 251 1945 J F Nye *Physical Properties of Crystals Their Representation by Tensors and Matrices* 1957

Pyrolusite

A mineral having composition MnO_2 Pyrolusite is the most important ore of manganese It crystallizes in the tetragonal system but well developed ... 2 on the Mohs scale ... the specific gravity is 4 75 Crystals of polianite show a perfect prismatic cleavage and have a hardness of 6 and a specific gravity of 5 1 The luster is metallic and the color iron black It frequently forms pseudomorphs after other manganese minerals notably manganite

Pyrolusite is a secondary mineral formed by the alteration of other manganese minerals such as manganite psilomelane rhodochrosite and rhodonite Manganese dissolved from rocks by surface solutions may be redeposited as pyrolusite as dendritic coatings on the walls of fractures as nodules on the sea bottom and as beds in residual clays Pyrolusite is extensively mined as a manganese ore in many countries chiefly in Russia Ghana India the Union of South Africa Morocco Brazil and Cuba The chief use of manganese is in making spiegeleisen and ferromanganese employed in steel manufacture It is also used as an oxidizer in the production of chlorine bromine and oxygen in electric cells and batteries and as a decolorizer in glass See FERROALLOY MANGANESE MANGANITE PSILOMELANE RHODOCHROSITE RHODONITE [C S HU]

Pyrolysis

The chemical transformation of a material into one or more new substances solely through the application of heat The new substance formed results only from rearrangement of atoms (or molecules) present in the original parent material If more than one parent material participates in the chemical reaction even though heat is employed the resulting chemical changes are not pyrolytic but of some other nature (for example combustion or hydrogenation)

Pyrolytic reactions may however occur in the presence of other compounds which do not themselves decompose and appear in the reaction products The use of solvents or catalysts is an example in this respect (see CATALYSIS) Use of γ radiation during pyrolytic decomposition is a similar case

Pyrolytic processes were among the first chemical reactions used by early chemists to study naturally occurring materials The production of oxygen by heating mercuric oxide and the pyrolytic decomposition of natural rubber to isoprene are ex

amples Wood alcohol (methanol) originally was obtained by pyrolyzing hardwood

Modern industry employs pyrolytic processes extensively The cracking of petroleum the production of carbon black the chemical manufacture of ethylene biphenyl and ketene and the carbonization of coal to coke coal chemicals and gas are examples Limestone is calcined (pyrolyzed) to make lime *See* CARBON BLACK CHARCOAL COAL CHEMICALS COKE CRACKING, DESTRUCTIVE DISTILLATION LIME (INDUSTRIAL) OIL SHALE

Most pyrolytic processes of industrial importance are carried out at high temperature Technically however the heat level must be sufficient only to impart the requisite thermal energy to break down the original chemical structure and induce chemical rearrangement The liquid vapor gas or solid phases are all employed Pressure although usually thermodynamically disadvantageous is often used to reduce gas volumes handled increase reaction rates or maintain a liquid phase Vacuum is also used Thus *n* butane may be pyrolyzed in a tubular heater to yield principally ethylene at 1415 1450°F and 22 psia Acetic acid is pyrolyzed industrially to ketene at 1300°F and at a pressure of only slightly over ¼ atm In the laboratory benzazide ($C_6H_5CON_3$) can on the other hand be converted by heat to phenyl isocyanate (C_6H_5NCO) at only 140°F and normal pressure in a benzene solvent

Approximately one half of all industrial organic chemicals in the United States are produced from either petroleum or natural gas (petrochemicals) and about one fourth from coal (coal chemicals) The principal process first employed to secure starting materials for further synthesis is pyrolysis Petroleum natural gas and coal are the initial raw materials

Petroleum Crude petroleum or a fraction thereof upon cracking (pyrolysis) yields various liquid fractions such as gasoline and kerosine and light gases including ethylene propylene butylenes and cyclohexanes The latter gases are the principal intermediates for further organic synthesis Thermal cracking (pyrolysis) is conducted at 900–1100°F and pressures of 600 to 1000 psi in liquid or mixed liquid vapor phase Catalytic cracking in the presence of aluminum silicates is a similar pyrolytic process but it is conducted at lower temperatures and pressures (850°F 35 psi) Coking or visbreaking is a related pyrolytic procedure

Chemically the cracking process involves the thermal decomposition of large molecules to smaller ones *Thus in a simplified example a* charge stock may be pyrolyzed to a heavy gas oil of 29 carbon atoms molecular size This in turn by further pyrolysis yields hydrocarbons of 5 to 8 carbon atoms *(gasoline) plus ethylene gas* The ethylene is then converted by further synthesis to such industrial organic chemicals as polyethylene ethyl alcohol ethylene glycol and ethyl ether In a similar way pyrolysis may yield propylene for conversion to butadiene (for synthetic rubber) and hexamethylenediamine (for nylon)

Natural gas The principal hydrocarbons separated from natural gas for further industrial organic synthesis are methane ethane propane and butane By pyrolytic processes these can be converted to compounds such as carbon black methanol formaldehyde chloroform and polypropylene Molecular rebuilding processes employed in this connection include isomerization (converting straight chains to branched chains) dehydrogenation (removing hydrogen to yield olefins) thermal polymerization (building large molecules from smaller ones) and aromatization (converting aliphatic to aromatic hydrocarbons) These four processes are all catalyzed pyrolytic reactions

Coal The pyrolysis of coal is carried out at either high temperature (1450–1850°F) or low temperature (950–1450°F) to yield coke benzene, naphthalene toluene xylenes and cresols as principal intermediates These are converted by further organic synthesis to substances such as aniline TNT phthalic anhydride dyes and plastics

The utilization of oil shale is a coming new industry The first step in treating oil shale is a pyrolytic decomposition of the organic matter present by retorting to yield shale oil and gases for further processing to synthetic fuels and organic chemicals *See* UNIT PROCESSES [C.H P]

Bibliography C D Hurd *Pyrolysis of Carbon Compounds* 1929 W L. Nelson *Petroleum Refinery Engineering* 4th ed 1958

Pyrometallurgy

Processes employing chemical reactions at elevated temperatures for the extraction of metals from ores and concentrates The use of heat to cause reduction of copper ores by charcoal dates from before 3000 B C The techniques of pyrometallurgy have been gradually perfected as knowledge of chemistry has grown and as sources of controlled heating and materials of construction for use at high temperature have become available Pyrometallurgy at mid twentieth century was the principal means of metal production

The advantages of high temperature for metallurgical processing are several chemical reaction rates are rapid reaction equilibria change so that processes impossible at low temperature become spontaneous at higher temperature and production of the metal as a liquid or a gas facilitates physical separation of metal from residue

The processes of pyrometallurgy may be divided into preparation processes which convert the raw material to a form suitable for further processing (for example roasting to convert sulfides to oxides) reduction processes which reduce metallic compounds to metal (the blast furnace which reduces iron oxide to pig iron) and refining processes which remove impurities from crude metal (fractional distillation to remove iron lead and cadmium from crude zinc)

The complete production scheme from ore to refined metal may employ pyrometallurgical processes (steel lead tin zinc) or only the primary extraction processes may be

other methods used for refining (copper, nickel) In some cases (uranium, tungsten molybdenum), isolated pyrometallurgical processes are used in a treatment scheme which is predominately non pyrometallurgical See IRON (EXTRACTION FROM ORE), METALLURGY, PYROMETALLURGY, NONFERROUS [H H K]

Pyrometallurgy, nonferrous

Methods for the extraction of the nonferrous metals from ores and concentrates based on chemical reactions at high temperatures Pyrometallurgy was the earliest method (around 3000 B C) by which man recovered metals from ore minerals and the only significant means of metal extraction until the late nineteenth century when electrometallurgy and hydrometallurgy were introduced (see ELECTROMETALLURGY, HYDROMETALLURGY) Improvements in nonferrous pyrometallurgy during the twentieth century have helped it to retain its preeminent position in the face of growing competi

other materials of construction the introduction of electrical control devices and electrical heating the design of vacuum and pressure processes the availability of low cost oxygen and many other engineering improvements have radically altered twentieth-century pyrometallurgy Some of the noteworthy new processes of nonferrous pyrometallurgy that date from the period 1925–1958 are the continuous vertical retort process for zinc smelting and the fractional distillation process for refining zinc the vacuum retort process (Pidgeon process) for production of magnesium and calcium the Kroll process for the production of titanium and zirconium flash smelting of low grade nickel concentrates in oxygen and the zinc blast furnace process

Table 1 summarizes production data for 10 major nonferrous metals Pyrometallurgy is still the major means for production of nonferrous metals, both in number of applications and in total tonnage of metal produced In addition to the metals listed in Table 1 pyrometallurgy has significant applications to the production of uranium cobalt silver mercury bismuth molybdenum beryllium tungsten zirconium gold tantalum niobium (columbium) and other minor metals The production of iron steel ferromanganese and ferrochrome is entirely pyrometallurgical See FERROALLOY IRON (EXTRACTION FROM ORE)

Several advantages result from the use of high temperature for extraction of metals from ores Chemical reaction rates increase as tempera

and can in some cases, be shifted to a preferred direction by increase in temperature. Thus the reduction of zinc oxide by carbon at atmospheric pressure is impossible because of equilibrium considerations below about 900°C, but is spontaneous at higher temperatures

The effect of temperature on the physical separation of the reduced metal from other products of the reaction is as important as its effect on reaction rate and equilibrium Reduction of ores and concentrates invariably results in two or more products—the reduced metal and a residue composed of substances such as unreduced gangue minerals (silicon dioxide SiO_2, aluminum oxide Al_2O_3, calcium oxide CaO, ferrous oxide, FeO), and the

ture processes offer several simple methods of making this separation provided that the products can be liquefied or selectively vaporized

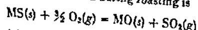
The processes of pyrometallurgy are discussed below under the three general types preparatory, reduction and refining processes

PREPARATORY PROCESSES

Preparatory processes convert the raw material (ore or concentrate) to a chemical form suitable for further processing A few of the more common processes are described below

Roasting of sulfides The burning of metallic sulfides in air or oxygen is used to convert sulfide minerals to metallic oxides, sulfates, or both together with sulfur dioxide for the following purposes (1) To prepare the material for reduction by carbon or carbon monoxide metallic oxides can be reduced by these agents but metallic sulfides cannot (see Table 2) (2) To prepare the material for matte smelting, there must be just enough sulfur in the feed to form a matte of desired composition with the valuable metals (matte is a molten solution of metallic sulfides) Excess sulfur is removed from the concentrate by incomplete roasting (3) To prepare the material for acid leaching and electrolysis, the feed must be metallic oxide because metallic sulfides are relatively insoluble in sulfuric acid the commonly used leaching agent (4) To manufacture sulfuric acid or sulfur dioxide, conversion of the sulfur dioxide in roaster gases to sulfuric acid or liquid sulfur dioxide is usually a legal necessity if not always a profitable enterprise

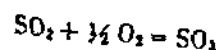
The principal reaction during roasting is



$$MS(s) + \tfrac{3}{2} O_2(g) = MO(s) + SO_2(g)$$

where (s) and (g) indicate solid and gas phases, respectively For all the common sulfides this reaction is strongly exothermic and spontaneous left to right provided the roaster gas contains a little free oxygen Most roasting operations are carried out at 650–1000°C

Roaster gases always contain some sulfur trioxide, formed by the reaction



$$SO_2 + \tfrac{1}{2} O_2 = SO_3$$

The equilibrium of this reaction shifts to the left as the temperature is raised so that roasting at

amples Wood alcohol (methanol) originally was obtained by pyrolyzing hardwood

Modern industry employs pyrolytic processes extensively The cracking of petroleum, the production of carbon black, the chemical manufacture of ethylene biphenyl, and ketene and the carbonization of coal to coke, coal chemicals, and gas are examples Limestone is calcined (pyrolyzed) to make lime *See* CARBON BLACK; CHARCOAL, COAL CHEMICALS, COKE, CRACKING, DESTRUCTIVE DISTILLATION, LIME (INDUSTRIAL), OIL SHALE

Most pyrolytic processes of industrial importance are carried out at high temperature Technically, however, the heat level must be sufficient only to impart the requisite thermal energy to break down the original chemical structure and induce chemical rearrangement The liquid, vapor, gas, or solid phases are all employed Pressure, although usually thermodynamically disadvantageous, is often used to reduce gas volumes handled increase reaction rates, or maintain a liquid phase Vacuum is also used Thus, *n* butane may be pyrolyzed in a tubular heater to yield principally ethylene, at 1415–1450°F and 22 psia Acetic acid is pyrolyzed industrially to ketene at 1300°F and at a pressure of only slightly over ¼ atm In the laboratory, benzazide ($C_6H_5CON_3$) can, on the other hand, be converted by heat to phenyl isocyanate (C_6H_5NCO) at only 140°F and normal pressure, in a benzene solvent

Approximately one-half of all industrial organic chemicals in the United States are produced from either petroleum or natural gas (petrochemicals), and about one fourth from coal (coal chemicals) The principal process first employed to secure starting materials for further synthesis is pyrolysis Petroleum, natural gas, and coal are the initial raw materials

Petroleum. Crude petroleum, or a fraction thereof, upon cracking (pyrolysis) yields various liquid fractions, such as gasoline and kerosine, and light gases including ethylene propylene, butylenes, and cyclohexanes The latter gases are the principal intermediates for further organic synthesis Thermal cracking (pyrolysis) is conducted at 900–1100°F and pressures of 600 to 1000 psi in liquid or mixed liquid vapor phase Catalytic cracking in the presence of aluminum silicates is a similar pyrolytic process but it is conducted at lower temperatures and pressures (850°F, 35 psi) Coking or visbreaking is a related pyrolytic procedure

Chemically, the cracking process involves the thermal decomposition of large molecules to smaller ones Thus, in a simplified example, a charge stock may be pyrolyzed to a heavy gas oil of 29 carbon atoms molecular size This in turn by further pyrolysis yields hydrocarbons of 5 to 8 carbon atoms (gasoline), plus ethylene gas The ethylene is then converted by further synthesis to such industrial organic chemicals as polyethylene, ethyl alcohol ethylene glycol and ethyl ether In a similar way, pyrolysis may yield propylene for conversion to butadiene (for synthetic rubber) and hexamethylenediamine (for nylon)

Natural gas. The principal hydrocarbons separated from natural gas for further industrial organic synthesis are methane, ethane, propane and butane By pyrolytic processes, these can be converted to compounds such as carbon black, methanol, formaldehyde, chloroform and polypropylene Molecular rebuilding processes employed in this connection include isomerization (converting straight chains to branched chains), dehydrogenation (removing hydrogen to yield olefins), thermal polymerization (building large molecules from smaller ones), and aromatization (converting aliphatic to aromatic hydrocarbons) These four processes are all catalyzed pyrolytic reactions

Coal The pyrolysis of coal is carried out at either high temperature (1450–1850°F) or low temperature (950–1450°F) to yield coke benzene naphthalene, toluene xylenes, and cresols as principal intermediates These are converted by further organic synthesis to substances such as aniline TNT, phthalic anhydride, dyes, and plastics

The utilization of oil shale is a coming new industry The first step in treating oil shale is a pyrolytic decomposition of the organic matter present, by retorting to yield shale oil and gases for further processing to synthetic fuels and organic chemicals *See* UNIT PROCESSES [C.H P]

Bibliography C. D Hurd, *Pyrolysis of Carbon Compounds*, 1929, W L. Nelson *Petroleum Refinery Engineering*, 4th ed , 1958

Pyrometallurgy

Processes employing chemical reactions at elevated temperatures for the extraction of metals from ores and concentrates The use of heat to cause reduction of copper ores by charcoal dates from before 3000 B C The techniques of pyrometallurgy have been gradually perfected as knowledge of chemistry has grown and as sources of controlled heating and materials of construction for use at high temperature have become available Pyrometallurgy at mid twentieth century was the principal means of metal production

The advantages of high temperature for metallurgical processing are several chemical reaction rates are rapid reaction equilibria change so that processes impossible at low temperature become spontaneous at higher temperature and production *of the metal as a liquid or a gas facilitates physical* separation of metal from residue

The processes of pyrometallurgy may be divided into preparation processes which convert the raw material to a form suitable for further processing (for example roasting to convert sulfides to oxides), reduction processes which reduce metallic compounds to metal (the blast furnace which reduces iron oxide to pig iron) and refining processes which remove impurities from crude metal (fractional distillation to remove iron lead and cadmium from crude zinc)

The complete production scheme from ore to refined metal, may employ pyrometallurgical processes (steel lead tin zinc) or only the prima traction processes may be llurg

pplications for roasting and other gas solid reac ons Air is pumped upward through the perforated earth at a rate sufficient to expand the bed of alcine particles and cause a mixing action known s fluidization Only very fine dust particles are car ied out by the gas stream since the gas velocity is elatively low above the bed Sulfide concentrate is ed continuously by means of the screw conveyor, and the calcine leaves by overflow through the pipe provided at the top of the fluid bed Circulation within the bed results in a remarkable homogeneity of temperature and gas composition from point to point Burning of the sulfides provides the heat required to maintain the temperature Temperature may be controlled by automatic injection of water into the bed The possibility of close control of temperature and gas composition in the bed and the absence of moving parts in the hot zone are distinct advantages of this roaster It is unsuitable for some extremely fine feeds (smaller than 100-mesh) or for feed particles which stick together (agglomerate) on heating

In the multiple hearth roaster the hot roaster gases pass over beds of concentrate held on refractory hearths In the flash roaster the pulverized concentrate is mixed with an air stream and injected into a hot combustion chamber

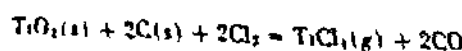
Sintering Finely divided solid particles are consolidated into relatively dense aggregates in pyrometallurgy to prepare lump feed for the blast furnace and other reduction processes The common sintering device is the downdraft (Dwight Lloyd) sintering machine shown in Fig 2 The feed must contain a proportion of solid fuel which may be sulfide minerals or fine coal if a nonsulfide material is to be roasted The feed is spread continuously as a bed (4–8 in deep) on the surface of a modified pan conveyor Air passes through the bed when the gratelike pans pass over a suction box located beneath the conveyor The top of the bed is ignited as it passes under a fuel fired ignition box The zone of ignition burns down through the feed bed in a manner similar to the burning of a cigar The local temperature in the burning zone is high enough to cause partial fusion of the products into a clinker or sinter cake Downdraft sintering is used as a desulfurizing process (roasting) as well as f r sintering in the treatment of lead sulfide concentrate

Chlorination Metal oxide ores and concentrates are converted to metal chlorides in the production of titanium zirconium and other refractory metals In the case of titanium and zirconium the need for chl rination arises from the fact that direct reduc t n of the oxides of these metals is difficult whereas the chlorides are readily red iced by either metallic sodium or magnesium (see TITANIUM)

In the case of titanium the chlorination reaction is

$$TiO_2(s) + 2C(s) + 2Cl_2 = TiCl_4(g) + 2CO$$

A red icing agent such as carbon is required to make the reaction spontaneous from left to right.



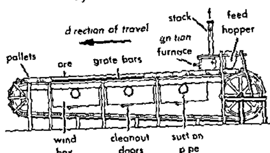


Fig 2 Cross section of Dw ght Lloyd downdraft s ntering machine (From R E Kirk and D F Othmer eds Encycloped a of Chemical Technology Interscience 1947–1956)

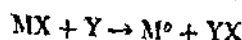
The reaction is slightly exothermic but the amount of heat generated may be insufficient to maintain the reaction temperature of 800–900°C

Chlorination may be carried out with fine particle feeds in a fluid bed similar to that shown in Fig 1 Alternatively the metal oxide and carbon may be sintered together or briqueted and then chlorinated in a shaft furnace When the metal chloride is formed as a liquid rather than a gas the shaft furnace is used and the liquid metal chloride is removed from the bottom of the shaft Internal electric heating has been used to supply the extra heat needed in shaft furnace chlorination

Drying and calcination Quite often free water is evaporated and hydrates and carbonates are decomposed Examples are the drying of wet sulfide concentrates in preparation for flash roasting the decomposition of aluminum trihydrate to alumina to prepare feed for the electrolytic production of aluminum the decomposition of dolomite ($CaCO_3$ $MgCO_3$) to calcined dolomite (CaO MgO) to prepare feed for pyrometallurgical production of magnesium metal (D d–
chines us
the fluid b
furnace (
all cases g and calcination are strongly endothermic processes so that heat must be supplied both to maintain the temperature and to satisfy the heat requirements of the reaction

REDUCTION PROCESSES

These processes



quires the use of a reducing agent a substance which will combine with the unwanted element in the metal compound If MX represents a metal compound and Y a reducing agent then the general reaction for reduction is

$$MX + Y \rightarrow M^{\circ} + YX$$

where M° indicates a free metal To achieve a spontaneous reaction from left to right YX must be a more stable compound than MX under the conditions of the process In more exact terms the free-energy change for the reaction must be negative

Table 1 Extraction and refining methods for 10 major nonferrous metals

Metal	U S consumption 1956, short tons	Extraction*	Refining*
Aluminum	2 055 000	electro	hydro†
Copper	1,521,000	90% pyro, 10% hydro-electro	90% electro 10% pyro
Lead	1,210,000	pyro	80% pyro, 20% electro
Zinc	1,009,000	60% pyro, 40% electro	60% pyro 40% electro
Sodium	135,000‡	electro	hydro†
Nickel	128,000	80% pyro, 20% hydro	electro
Tin	66 800	pyro	pyro
Magnesium	53 000	95% electro, 5% pyro	hydro†
Antimony	12 900	pyro	pyro
Titanium	10,900	pyro	pyro†

* Pyro pyrometallurgy, electro electrometallurgy, hydro, hydrometallurgy, all percentages are estimates.
† The metal compound is refined prior to extraction ‡ Estimated

1000°C produces very little SO_3 Low temperature and corresponding high concentration of SO_3 in the roaster gas favors the formation of sulfates rather than oxides

The most widely used roasting furnaces are of three types multiple-hearth fluid bed and flash roaster The feed to the roaster is pulverized (10-mesh or finer) and the solid product (calcine) is also fine Roaster gases carry much dust which is collected and returned to the primary calcine by standard dust-collecting equipment.

A fluid bed roaster is illustrated in Fig 1 This device was introduced into pyrometallurgical practice after World War II and has found numerous

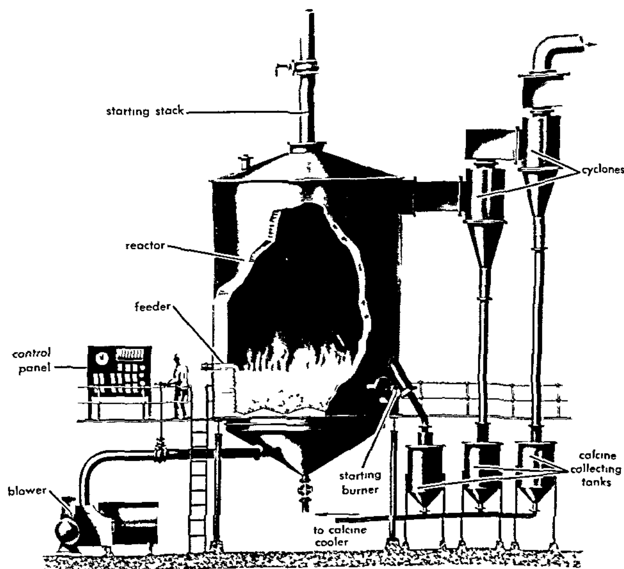
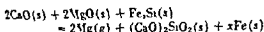


Fig 1 Fluid bed roaster and auxiliary equipment (*Dorr-Oliver, Inc*)

production of magnesium and calcium It was widely used during the war for the production of magnesium Only a small production of Mg^0 and Ca^0 by this process remained in 1958 The reaction in the case of Mg^0, is



$$2CaO(s) + 2MgO(s) + Fe_xSi(s) = 2Mg(g) + (CaO)_2SiO_2(s) + xFe(s)$$

Silicon in the relatively inexpensive form of ferrosilicon is the reducing agent The raw material is calcined dolomite. At 1200°C, the equilibrium pressure of $Mg(g)$ in this reaction is only 34 mm Hg For this reason and also to permit condensation of massive solid Mg^0 in the condenser, the reaction is carried out in a high vacuum

Retorts must be capable of withstanding an atmosphere of pressure difference at 1200°C and are made from special Cr Ni Fe alloys Like the Belgian retort they are small (10-in inside diameter by 10-ft length) are placed horizontally in a furnace containing many retorts and are operated on a batch basis The condenser consists of a removable sleeve that fits into the cool end of the retort which protrudes from the furnace The disadvantages of the process are the high cost of ferrosilicon and maintenance of retorts The advantages are a much simplified process and purer product compared with the Dow sea water process (*see* MAGNESIUM)

Reduction to liquid metal. This is the most common form of metal reduction process, and it is used for the involatile metals of moderate melting point Cu^0, Pb^0, Sn^0, Ni^0, Sb^0, Ag^0, Co^0, Be^0, Bi^0, and U^0 If the metal alone is liquefied, some separation of metal from solid residue is possible by liquation the draining of liquid metal away from the reaction mass It is more common to add substances (fluxes) which form a second liquid phase (slag) with the residue, thus permitting a more complete and simplified separation of the liquid metal The common furnaces for production of liquid metal are the blast furnace (shaft furnace), reverberatory furnace, and electric arc furnace.

The cross section of a lead blast furnace is shown in Fig 3 It consists of a shaft of rectangular cross section, 20–28 ft high, 15–22 ft long and tapering in width from 6–10 ft at the top to 4–5 ft at the bottom The feed added intermittently at the top of the shaft consists principally of sinter cake (roasted and sintered lead concentrates plus fluxes) and lump coke Air is blown into the furnace

Fig 4 Longitudinal and cross section of reverberatory furnace for smelting copper (Harbison-Walker Refractories Co)

Table 2 Reducing agents in nonferrous pyrometallurgy

Reducing agent	Approx cost		Reducing strength for†			Application to production of
	¢/lb	¢/g-equiv*	oxides	sulfides	halides	
Carbon (coke)	0.7	0.0093	S	N	N	Zn^0, Ni^0 Co^0, Sn^0
CO (from coke)	0.3	0.0093	M	N	N	Pb^0 Cu^0, Sn^0
S (from MS)	‡	‡	W	N	N	Cu^0, Pb^0, Hg^0
H_2 (from natural gas)	9.0	0.02	M	W	M	Mo^0, W^0
Fe^0 (scrap)	2.8	0.17	M	M	M	Cu^0 Pb^0, Sb^0
Si^0 (ferrosilicon)	15.0	0.23	S	M	M	Mg^0
Al^0	26.8	0.53	VS	M	S	Ca^0
Na^0	17.0	0.86	S	S	VS	Ti^0, Zr^0
Mg^0	35.3	0.94	VS	S	VS	Ti^0, Zr^0, Hf^0 Be^0 U^0

* This gives a more valid comparison of the relative cost of different reducing agents

† VS very strong, S strong M moderately strong, W, weak, N little or none

‡ The value of S in a sulfide concentrate is uncertain but very small

This condition may be obtained by suitable choice of Y, by adjustment of the concentrations of the reactants and products, and by the choice of temperature and pressure for the process

Other considerations in the selection of a reducing agent are that (1) it should be inexpensive with respect to the value of the metal reduced, and (2) *the product YX should be in a form which can* be readily separated from the metal, M^0 The common reducing agents of pyrometallurgy are listed in Table 2, with their approximate costs and the types of reduction for which they are suitable

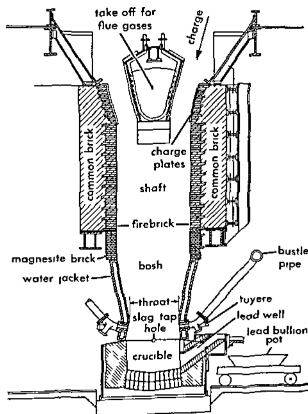
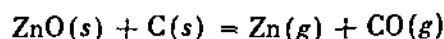


Fig 3 Cross section of lead blast furnace (*From R E Kirk and D E Othmer, eds, Encyclopedia of Chemical Technology*, Interscience 1947–1956)

Reduction processes are best characterized by the physical form of the metal resulting—gaseous liquid, or solid

Reduction to gaseous metal. Only the volatile metals, zinc, Zn, cadmium, Cd, mercury, Hg, alkali, and alkaline earth metals may be processed in this way Industrial practice exists for production of Zn^0, Hg^0, Mg^0 (magnesium), and Ca^0 (calcium) The principal advantage of this method lies in the simplicity and completeness of the separation of reduced metal from residue—the gaseous metal can be condensed to liquid or solid in a condenser physically separated from the reactants and residue

Zinc (boiling point 906°C) is produced by reduction of zinc oxide, ZnO with carbon The reaction



$$ZnO(s) + C(s) = Zn(g) + CO(g)$$

is carried out at 1200–1300°C in retorts from which air is excluded The Belgian retort is still widely used, though it is a small batch process which produces only 40–60 lb of zinc per day per retort The retorts and condensers are made from fire clay They are fixed in an almost horizontal position in a fuel fired furnace which may hold several hundred retorts Much hand labor is required for filling the retorts with reactants, removing residue and emptying condensers though mechanical devices have been developed for some of these purposes Poor condensation efficiency results from the difficulty of excluding air from the condenser

Continuous, vertical retorts have been developed Some are fuel fired their capacity is 5 tons of zinc per day per retort Others are electric arc heated, their capacity is 25 tons of zinc per day per retort These processes are generally superior to the Belgian retort in metallurgical efficiency and labor requirements per ton of zinc produced but high capital cost of the vertical retort processes and special economic considerations render the Belgian retort process competitive in certain localities

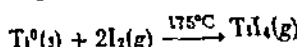
The vacuum retort process was developed by L M Pidgeon just prior to World War II for the

erized by the kind of separation employed—gas from liquid or solid (volatilization), solid from liq ... or one liquid from a

...

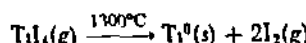
he base metal may be removed as a gas Crude zinc (containing small amounts of Fe and Pb) is refined by fractional distillation where the *more volatile* base metal (zinc) is distilled away from a residue of lead and iron Crude lead (containing a small amount of zinc) is refined by batch *vacuum distillation* where only the impurity (zinc) is volatilized *See* DISTILLATION

The iodide process (or van Arkel–de Boer process) is a volatilization process applicable to the involatile metals As adapted to titanium iodine vapor is passed in contact with the crude solid metal at low temperature (175°C) and the titanium reacts to form gaseous TiI_4



$$Ti^0(s) + 2I_2(g) \xrightarrow{175°C} TiI_4(g)$$

Most impurities remain as a solid residue The TiI_4 vapor is transported to an electrically heated wire (1300–1400°C) where the compound is decomposed to pure solid Ti^0 The iodine gas is regenerated and reused in a cyclic process



$$TiI_4(g) \xrightarrow{1300°C} Ti^0(s) + 2I_2(g)$$

This process has been widely used for preparing highly purified Ti^0 Zr^0 Hf^0 Si^0 and other metals

Drossing This process depends on the change of solubility of an impurity in the liquid base metal as the temperature is changed When on cooling a crude liquid metal the impurity precipitates as a solid the precipitated solid may be removed from the liquid by scraping it off the surface of the liquid (drossing) if the solid is lighter than the liquid or by filtration Iron is removed as a solid from crude tin and copper is removed as a solid from crude lead in this manner

Refining by precipitation is similar to drossing except that a precipitating agent is added to the crude liquid metal to precipitate the impurity as a solid compound Sulfur may be added to liquid lead to precipitate solid copper sulfide A very important application of precipitation is the Parkes process in which zinc is added to crude liquid lead to precipitate solid intermetallic compounds of zinc with gold and silver By this means the small but valuable amounts of silver and gold found in most crude lead can be concentrated and recovered profitably

Slag refining In this process crude liquid metal is heated in contact with a second immiscible liquid (slag or molten salt) and the impurities are absorbed into the second liquid Success of the process depends on the selective oxidation of impurities and the solution of the oxidized impurities in the slag or molten salt For this reason the process is useful only for removing impurities which are more easily oxidized than the base metal As an example in the slag refining of copper air is blown into the crude liquid copper and a slag composed mainly of Cu_2O is formed This slag absorbs many impurities from the metal (oxides of Fe, Pb Zn Sn As and Sb) Impurities which are less easily oxidized than the copper (gold silver) are not removed Slag refining is usually conducted in reverberatory furnaces It is used in the refining of copper, lead silver and other metals *See* METALLURGY, UNIT OPERATIONS [H H K]

Bibliography J L. Bray, *Non ferrous Production Metallurgy*, 2d ed , 1947, C R Hayward, *An Outline of Metallurgical Practice*, 3d ed , 1952, D M Liddell (ed) *Handbook of Non ferrous Metallurgy*, 2d ed , 2 vols 1945

Pyrometer

A temperature measuring device commonly applied to instruments that measure high temperatures Actually some pyrometers are used in the same temperature range as thermometers, but others are suitable only for higher temperatures This article discusses radiation and optical pyrometers used in radiation pyrometry For other temperature measuring devices *see* THERMOCOUPLE, THERMOMETER

Radiation pyrometry is the measurement of the temperature of an object by measuring some characteristic of the energy it radiates All objects radiate energy but the total radiation the radiation at every wavelength and the location of the radiant energy maximum vary with the temperature of the body The radiation also varies with the size of the object and the character of its surface A black body by definition reflects zero radiant energy and emits the maximum energy at any given temperature The black body is an idealized physical concept approached closely only by certain geometric arrangements incorporating multiple reflections from highly absorptive surfaces However it is useful for theoretical and comparative purposes The emissivity of a body is the ratio of its actual radiant energy to what it would be if it were truly black

Radiation received from a body varies as the difference between the fourth power of its absolute temperature and the fourth power of the absolute temperature of the body receiving the radiant energy Radiation pyrometers intercept and measure a small but fixed portion of the radiation from an object to determine its temperature (Fig 1)

Radiation pyrometers These are designed to pick up a broad spectrum of the radiant energy and concentrate it on a highly responsive thermal ele

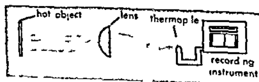


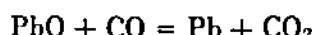
Fig 1 Elementary rad ation pyrometer (From D M Cons d ne, ed , *Process Instruments and Controls Handbook* McGraw H ll 1957)

through water cooled nozzles (tuyeres) located along each side of the furnace near the bottom The coke burns near the tuyeres producing both heat and CO, *the reducing agent*



$$C + \tfrac{1}{2}O_2 = CO \qquad \Delta H = -26\,400 \text{ cal/mole}$$

where ΔH is the heat of reaction The hot gases rise through the lump charge causing reduction of lead oxide



$$PbO + CO = Pb + CO_2$$

and transfer of heat to the solids The molten lead (melting point 327°C) trickles down through the charge and enters the crucible below the tuyeres The slag (principal constituents FeO, SiO_2, and CaO) does not liquefy until it reaches the zone near the tuyeres where the temperature is highest The liquid slag (density 3 5–4 0 g/cm³) floats on top of the lead (density 11 g/cm³) and is removed periodically through a tap hole in the furnace wall The molten lead (bullion) is removed by overflow through the lead well Lead blast furnaces have capacities of 100–400 tons of lead per day Blast furnaces for smelting of copper and tin are different in detail, but similar in principle to the lead blast furnace

The reverberatory furnace (see Fig 4) may be used for metal reduction when solid coke or coal is the reducing agent Heat is supplied by burning of fuel (gas, oil, or powdered coal) in the space between the charge and the roof Heat is transferred to the charge by direct radiation from the flame and by reflection from the arched refractory roof

Fine particle feed may be used in the reverberatory furnace since the gas velocities are relatively low The reducing agent (usually coke) is mixed with the feed and added through ports in the side or top of the furnace Slag and metal are removed through tap holes located at appropriate levels in the side of the furnace The furnace may be used either as a batch or continuous process Sizes range from small furnaces holding one ton of metal to large ones holding several hundred tons

In the electric arc furnace the electric arc replaces the burning of fuel as a source of heat It is a more versatile furnace than the reverberatory, since the furnace atmosphere may be made either reducing or oxidizing in nature The reverberatory atmosphere must be somewhat oxidizing in order to burn the fuel

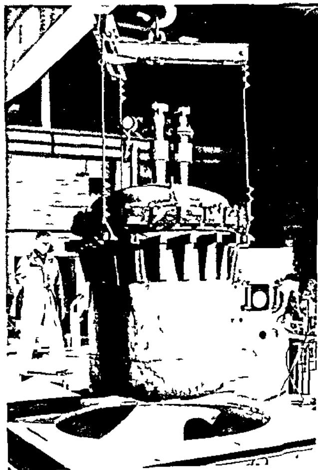
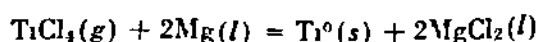


Fig 5 Reaction vessel for production of titanium being lowered into furnace (*Titanium Metals Corp of America*)

Reduction to solid metal. This process is employed when the melting point (mp) of the metal is unusually high (W^0, mp 3400°C, Mo^0, mp 2620°C) or when refractories to hold the liquid metal cannot be found (Ti^0, Zr^0) Compared with other reduction processes it is expensive cumbersome and usually a batch process The production of titanium by the Kroll process illustrates the problems involved

No refractory is known which will contain liquid titanium without reacting with the metal and rendering it impure, hence the metal must be produced as a solid at lower temperatures In the Kroll process a steel reaction vessel (Fig 5) is partly filled with ingots of magnesium and then sealed with a gastight cover The vessel is flushed with pure argon or helium because titanium reacts with and is rendered impure by all the common gases N_2 O CO CO_2 The vessel is then heated to about 750°C in a furnace When the magnesium has melted pure liquid $TiCl_4$ is fed slowly to the vessel through a tube passing through the cover The reduction reaction, which generates much heat is



$$TiCl_4(g) + 2Mg(l) = Ti^0(s) + 2MgCl_2(l)$$

where (l) designates the liquid phase

When the magnesium has been consumed much of the liquid $MgCl_2$ is drained through a tap hole in the bottom of the vessel which is then cooled to room temperature The cover is opened and the reaction mass interlocked crystals of Ti^0 plus $MgCl_2$ and some Mg^0, is removed by a *boring machine* The $MgCl_2$ and Mg^0 are then separated from the titanium by vacuum distillation or by leaching with dilute acid About 1500–3000 lb of titanium is produced in one batch by this process

REFINING PROCESSES

By these processes impurities are removed from crude metal to yield a product meeting market specifications Refining processes may be c c

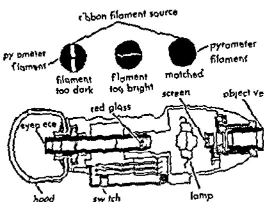


Fig 5 Optical pyrometer (Leeds and Northrup Co)

by adjusting its heating current in others the brightness of the image of the target is varied with

m ter measuring the heating current is directly cal brated in temperature units The melting tem peratures of gold (1336°K) palladium (1825°K) and platin im (2042°K) and the temperature of the carbon arc (about 3820°K) are used for stand ards Higher temperature calibrations are possible with absorbing glass screens Normal useful tem perature ranges lie between 1400 and 5200°F but the special techniques can extend the range to 10 000°F

The optical pyrometer is subject to errors due to reflect ons absorbing media in the optical path (gas glass or dirt) and nonblack body effects With care a well designed instrument is reported to produce an accuracy of ±6°F at 2000°F Opti cal pyrometers are temperature indicators and can not read ly be used for recording or control

Bolometer The bolometer is a Wheatstone bridge the four arms of which are thin grids of plat num (l The bridge is useful as a differential temperature device in radiation pyrometry Two opposite arms are subjected to the radiant en ergy and the other two are shielded from it See BOLOMETER see also TEMPERATURE MEASUREMENT [R E CL]

Bibliography D M Considine (ed) *Process Instruments and Controls Handbook* 1957

Pyrometric cone

One of a numbered series of ceramic compositions f rmed into triangular pyramids of height about 4 times the base and des gned so that each will sof ten and bend under its own weight after a certain heat treatment

The cones are used to indicate when ceramic ware has been adequately fired to check on tem perature uniform ty in a kiln and to determine the pyrometric cone equivalent of refractories See REFRACTORY

Cones are numbered from 022 021 020 the most eas ly fused thr ugh 02 01 1 2 to 42 the most re too close together some have been omitted (for ex ample 21 22 24 25) and where they were too widely spaced extra cones (31½ 32½) have been added

A cone is said to have reached its end point when its tip is bent to the level of the base for a constant rate of temperature rise this occurs at a definite temperature However the end point in dicates a heat treatment depending on both time and temperature the higher the rate of heating the higher the equivalent temperature of a given cone For example cone 022 is equivalent to 585°C (1085°F) when heated at 60°C (108°F) per hour and equivalent to 600°C (1112°F) when heated at 150°C (270°F) per hour

Proper firing of ceramic materials also requires a certain heat treatment therefore the end point of a given cone can be used to specify the firing of a ceramic ware (for example fired to cone 12) various time temperature schedules can be used to reach this point but the ware will be correctly fired Cones are most useful for firing ware made of materials similar to those in the cones (clay feld spar quartz) in other words the classical clay products For radically different materials such as the magnetic ceramic ferrites, cones are less useful as indicators of maturity

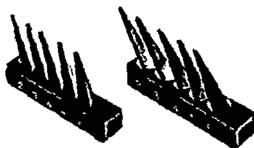


Fig 1 Pyrometric cones before firing showing two methods of setting in the cone plaque (Edward Orton Jr Ceramic Foundation)



Fig 2 Pyrometric cones after heat treatment These cones have been fired to cone 4 (Edward Orton Jr Ceramic Foundation)

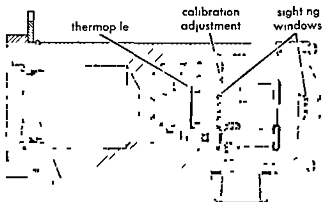


Fig 2 Radiation pyrometer (*Brown Instruments Div Minneapolis Honeywell Regulator Co*)

ment The range of the instrument determines the lens material in the optical system Calcium fluoride is used up to 1200°F, fused silica at 1000–2300°F, and Pyrex glass above 1500°F Other materials are available The thermal element also varies with the range and service and may be a photocell thermopile thermistor, thermocouple, resistance thermometer, or bolometer Since the thermal element measures a dynamic condition above ambient temperature within the pyrometer, and since this temperature is changed with the absorption of energy by the element the accuracy and speed of response of the pyrometer varies greatly with the element used and its associated compensating circuitry The electrical measuring system is usually a modified potentiometer (voltage measurement) or a modified Wheatstone bridge (resistance measurement)

The high temperature pyrometer (Fig 2), with various lenses is useful for temperature ranges above 400°F Since the energy received by any pyrometer falls rapidly at the low end of the range the scale is congested in this area and it is good practice to select an instrument so that the measured temperatures will be above the center of the range The low temperature pyrometer (Fig 3) is used for target temperatures between 125 and 700°F Because of the extremely low level of the radiant energy it incorporates a pyrometer hous

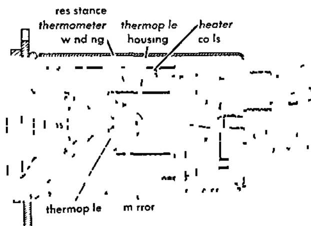


Fig 3 Low temperature radiation pyrometer (*Brown Instruments Div Minneapolis Honeywell Regulator Co*)

ing heater to minimize housing temperature variations

Radiation pyrometers are particularly useful for (1) temperatures above the practical operating ranges of thermocouples (2) environments which contaminate and limit the life of thermocouples (3) moving targets (4) targets not easily accessible, (5) targets which would be damaged by contact or insertion and (6) average temperatures of large surface areas

The accuracy of a pyrometer depends upon its range, its calibration and its use There are many unique sources of error including ambient temperatures, nonblack body radiation radiation absorption by intervening media (see Fig 4), and reflections Nevertheless the radiation pyrometer is a most useful device because it always indicates tem

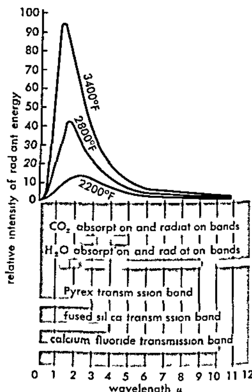


Fig 4 Gas radiation versus optical material transmission (*From D M Considine ed Process Instruments and Controls Handbook McGraw Hill 1957*)

perature variations and at high temperatures it is extremely sensitive Acceptable accuracies can be obtained by periodic calibration against other standards

Optical pyrometer This instrument shown in Fig 5 measures the intensity of a narrow band of radiant energy in the visible spectrum The temperature of the target is determined by an optical comparison of the relative brightness of the target with a source of known brightness such as a tungsten filament The instrument incorporates a red glass color filter so that the brightness comparison is made with wavelengths in the neighborhood of 0 65 micron In some designs the brightness of the standard is varied to match the image of the target

Pyroxene

A large group of common rock forming inosilicate (metasilicate) minerals (*see* SILICATE MINERALS) The extensive solid solution permitted between a variety of end members results in a complex com position for most of the pyroxenes The commonly recognized end members are listed below

Orthorhombic pyroxenes

Enstatite	$MgSiO_3$
Orthoferrosilite	$FeSiO_3$

Monoclinic pyroxenes

Diopside	$CaMg(SiO_3)_2$
Hedenbergite	$CaFe(SiO_3)_2$
Johannsenite	$CaMn(SiO_3)_2$
Jadeite	$NaAl(SiO_3)_2$
Acmite	$NaFe(SiO_3)_2$
Spodumene	$LiAl(SiO_3)_2$

Compositional relationships Solid solution be tween enstatite and orthoferrosilite and between clinoenstatite and clinoferrosilite is possible up to about 90 mole % of the iron end member Pure orthoferrosilite and clinoferrosilite are unknown, their compositions are represented by *iron olivine* (fayalite) and silica The intermediate composi tions of the enstatite series are commonly called hypersthene and the more iron rich compositions bronzite Solid solution between the various cal cium pyroxenes seems to be *complete* There is a limited solid solution permitted between the cal cium pyroxene series and the clinoenstatite series These minerals are generally referred to as augite There is a partial solid solution permitted between the clinoenstatite series and calcium pyroxene se ries these are called pigeonite *In addition* augite and ...

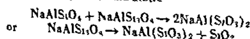
... ist in equilibrium in the mineral assemblage Solid solution between *diopside* and johannsenite is called schefferite zinc bearing *schefferite* is known as jeffersonite and solid solution between acmite and the calcium pyroxenes is known as *aegirinaugite* *Little* is known regarding the extent of solid solution between *acmite* and jadeite be tween jadeite and the other pyroxenes and between spodumene and the other pyroxenes *See* AUGITE ENSTATITE JADEITE PIGEONITE.

Phase relations The phase relationships be tween enstatite and clinoenstatite are *complicated* It was formerly held that clinoenstatite *represented* the high temperature modification of enstatite and inverted to enstatite on cooling It is now *known* that the true high temperature form of enstatite is also orthorhombic and is called protoenstatite which upon cooling inverts to enstatite but on rapid cooling inverts to clinoenstatite as a metastable form However if sufficient calcium is present cli noenstatite is apparently a stable form The effect of iron on the system is unknown *as yet* The amount of calcium pyroxene permitted in solid so lution with the enstatite series increases with in creasing temperature Therefore with decreasing temperature the calcium pyroxene becomes un stable and exsolves from the mineral along certain crystallographic directions in the form of thin plates—a very common feature in the enstatite *series of minerals*. Exsolution of the enstatite *orthoferrosilite* group from the calcium pyroxenes with decreasing temperatures is often absent in the mineral indicating either greater relative stability of the solid solution or greater difficulty in the ex solution process

Physical properties Microscopic work and often chemical analysis are needed to determine the chem ical composition although a few of the physical features suggest the dominance of certain end members for certain compositions The pyroxenes are characterized by two directions of well devel oped prismatic cleavages intersecting at approxi mately 87°, which distinguishes them from the related and similar appearing amphiboles (*see* AMPHIBOLE) The iron free pyroxenes are usually light in color (whitish gray or light green) with the color darkening to dark greens browns and blacks *with increasing* iron content The sodium pyroxenes are usually needlelike in crystal form whereas the other pyroxenes tend to be short *stubby* crystals The enstatite series have a wood like appearance and often a submetallic bronze luster probably the result of exsolutions and in clusions which seems to be absent from the cal cium pyroxenes Spodumene is most easily identi *fied* by its manner of occurrence and seems to be *restricted to* lithium rich pegmatites Spodumene is *generally* white although an emerald green vari ety (hiddenite) and a lilac variety (kunzite) are found *See* SPODUMENE

Occurrence The pyroxenes as a group are stable over a wide range of temperatures and occur as *major* and minor constituents in many volcanic igneous and metamorphic rocks Pyroxenes are major constituents in basalts gabbros norites py roxenites and many basic dikes Diopside can be ...

... most part in the highest grade *of* metamorphism (granulite facies) An aluminous diopside *called omphacite* is a characteristic mineral in the eclogite facies of *metamorphism* Pyroxenes are *found* as gangue minerals in certain ore deposits and occur in many slags and *meteorites* *Acmite* and jadeite are asso ciated with sodium rich *rocks* In addition jadeite is thought *to* represent the high pressure transfor mation of nephaline and albite



$$NaAlSiO_4 + NaAlSi_3O_8 \rightarrow 2NaAl(SiO_3)_2$$

or

$$NaAlSi_3O_8 \rightarrow NaAl(SiO_3)_2 + SiO_2$$

See DIOPSIDE ...

Crystal s... ships *of the* ... point of view ... hedral group ... oxygen atoms about a cen tral 4-valent silicon atom to form the SiO_4 anion results *in* an excess negative charge *for the* group

Cones are designed with end points roughly 20°C (36°F) apart; thus cones set in different parts of a kiln are a good check on temperature uniformity Even in automatically controlled kilns it is common practice to use cones as a means of checking the firing

Care should be exercised in using cones, since they are affected by factors other than time and temperature, such as the composition of the furnace atmosphere *See* CERAMIC TECHNOLOGY, KILN, SINTERING [M C M]

Bibliography Edward Orton Jr Ceramic Foundation, *The Properties and Uses of Pyrometric Cones,* 1951

Pyromorphite

A mineral series in the apatite group, or in the larger grouping of phosphate arsenate, and vanadate type minerals (*see* APATITE) In this series lead (Pb) substitutes for calcium (Ca) of the apatite formula $Ca_5(PO_4)_3(F\ OH\ Cl)$, and little fluorine (F) or hydroxide (OH) is present Substitution between phosphorus P and arsenic, As gives rise to a complete mineral series with pyromorphite $Pb_5(PO_4)_3Cl$ and mimetite $Pb_5(AsO_4)Cl$, as pure end members Substitution between As and vanadium V produces another series between mimetite and vanadinite $Pb_5(VO_4)_3Cl$

Pyromorphite generally is a chloride phosphate arsenate of lead $Pb_5(PO_4\ AsO_4)_3Cl$ The half of the series in which P > As is called pyromorphite, the remainder of the series with As > P is termed mimetite Vanadinite is a chloride vanadate of lead P and As may replace V in amounts up to P V = 1 4 7 and As V = 1 1

The pyromorphite series crystallizes in the hexagonal system Crystals are prismatic, pyramidal faces may be present Other forms are granular, globular, botryoidal Pyromorphite colors range through green yellow, brown, vanadinite occurs in shades of yellow, brown, red

Pyromorphites are widely distributed as secondary minerals in oxidized lead deposits. Pyromorphite is a minor ore of lead, vanadinite is a source of vanadium and minor ore of lead *See* LEAD, VANADIUM [W R L O]

Pyrotechnics

Although pyrotechnics or fireworks were probably known in the Orient at an earlier date they originated for the western world with the Greek fire of Byzantium in the seventh century A D At that time someone discovered the marvelous effect of saltpeter (KNO_3) on a fire and so launched the developments that centuries later led to guns and modern warfare Fireworks immediately became the fascinating playthings they are today and magnificent displays highlighted feasts and celebrations

The ingredients of fireworks are basically the familiar incendiary components Chlorates and nitrates are the oxidizers and sulfur, charcoal antimony sulfide and sometimes powdered metals are the fuels Colors are produced by the metals present in the salts sodium gives yellow, calcium brick red, strontium scarlet, barium, green and copper blue green Iron filings and aluminum powder give sparks and brilliant effusions The compounding and blending of these materials the combining of colors and effects and the patterning of these in time and space require great skill Secrets were handed from father to son, and every famous pyrotechnist had his specialty The Ruggieri family of Italy staged displays before many royal personages, and their fame spanned more than a century

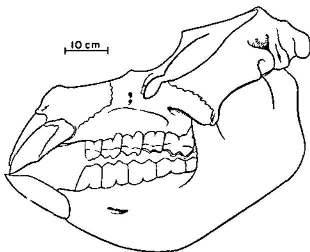
Law in many states requires that trains and trucks stopped on the road at night should burn a warning flare Such flares are usually made with strontium nitrate to impart a brilliant scarlet hue and employ a slow burning fuel of wax and saw dust

Colored smokes are used for military signaling and for daylight displays They incorporate a variety of vaporizable organic dyes in the usual pyrotechnic compositions *See* FUSE EXPLOSIVE, INCENDIARY, SMOKE [W E C]

Bibliography T L. Davis, *Chemistry of Powder and Explosives,* 1943

Pyrotheria

An extinct order of primitive mastodonlike, herbivorous hoofed mammals restricted to the Eocene and Oligocene deposits of South America There is only one family (Pyrotheriidae) in the order and four genera (*Pyrotherium, Propyrotherium Carlozittelia,* and *Griphodon*) in the family



Skull and jaw of Pyrotherium sorondi an early Oligocene pyrothere from South America (After F Loomis 1914)

The characters of this group superficially resembling those in early proboscideans are nasal openings over orbits indicating the presence of a trunk strong neck musculature and six upper and four lower bilophodont cheek teeth They also had four upper and two lower chisel like incisor tusks Pyrotheres are distantly related to the members of the superorder Paenungulata including Proboscidea Xenungulata and others *See* PROBOSCIDA XENUNGULATA [C T

Pyroxene

A large group of common rock forming *inosilicate* (metasilicate) minerals (*see* SILICATE MINERALS) The extensive solid solution permitted between a variety of end members results in a complex composition for most of the pyroxenes The commonly recognized end members are listed below

Orthorhombic pyroxenes

Enstatite	$MgSiO_3$
Orthoferrosilite	$FeSiO_3$

Monoclinic pyroxenes

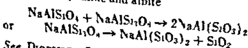
Diopside	$CaMg(SiO_3)_2$
Hedenbergite	$CaFe(SiO_3)_2$
Johannsenite	$CaMn(SiO_3)_2$
Jadeite	$NaAl(SiO_3)_2$
Acmite	$NaFe(SiO_3)_2$
Spodumene	$LiAl(SiO_3)_2$

Compositional relationships Solid solution between enstatite and orthoferrosilite and between clinoenstatite and clinoferrosilite is possible up to about 90 mole % of the iron end member Pure orthoferrosilite and clinoferrosilite are unknown, their compositions are represented by iron olivine (fayalite) and silica The intermediate compositions of the enstatite series are commonly called hypersthene and the more iron rich compositions bronzite Solid solution between the various calcium pyroxenes seems to be complete There is a limited solid solution permitted between the calcium pyroxene series and the clinoenstatite series *These* minerals are generally referred to as augite There is a partial solid solution permitted between the clinoenstatite series and calcium pyroxene series these are called pigeonite In addition augite and pigeonite ordinarily have aluminum substituting for both the silicon and iron magnesium atoms Consequently two monoclinic pyroxenes can coexist in equilibrium in the mineral assemblage Solid solution between diopside and johannsenite is called schefferite zinc bearing schefferite is known as jeffersonite and solid solution between acmite and the calcium pyroxenes is known as aegerinaugite Little is known regarding the extent of solid solution between acmite and jadeite between jadeite and the other pyroxenes and between spodumene and the other pyroxenes *See* ACMITE ENSTATITE JADEITE PIGEONITE

Phase relations The phase relationships between enstatite and clinoenstatite are complicated It was formerly held that clinoenstatite represented the high temperature modification of enstatite and inverted to enstatite on cooling It is now known that the true *high temperature* form of enstatite is also orthorhombic and is called protoenstatite which upon cooling inverts to enstatite but on rapid cooling inverts to clinoenstatite as a metastable form However if sufficient calcium is present, clinoenstatite is apparently a stable form The effect of iron on the system is unknown as yet The amount of calcium pyroxene permitted in solid solution with the enstatite series increases with increasing temperature Therefore, with decreasing temperature, the calcium pyroxene becomes unstable and exsolves from the mineral along certain crystallographic directions in the form of thin plates—a very common feature in the enstatite *series of minerals* Exsolution of the enstatite *orthoferrosilite* group from the calcium pyroxenes with decreasing temperatures is often absent in the mineral indicating either greater relative stability of the solid solution or greater difficulty in the exsolution process

Physical properties Microscopic work and often chemical analysis are needed to determine the chemical composition although a few of the physical features suggest the dominance of certain end members for certain compositions The pyroxenes are characterized by two directions of well developed prismatic cleavages intersecting at approximately 87°, which distinguishes them from the related and similar appearing amphiboles (*see* AMPHIBOLE) The iron free pyroxenes are usually *light in color* (whitish gray or light green) with *the color darkening* to dark greens browns and blacks with increasing iron content The sodium pyroxenes are usually needlelike in crystal form whereas the other pyroxenes tend to be short *stubby crystals* The enstatite series have a wood *like appearance* and often a submetallic bronze luster probably the result of exsolutions and inclusions which seems to be absent from the calcium pyroxenes Spodumene is most easily identified *by its* manner of occurrence and seems to be *restricted to lithium* rich pegmatites Spodumene is generally white although an emerald green variety (hiddenite) and a lilac variety (kunzite) are found *See* SPODUMENE

Occurrence The pyroxenes as a group are stable over a *wide* range of temperatures and occur as major and minor constituents in many volcanic igneous and metamorphic rocks *Pyroxenes* are major constituents in basalts gabbros norites pyroxenites and many basic dikes Diopside can be a common constituent in all but the lowest grade of metamorphic rocks The orthorhombic pyroxenes in metamorphic rocks are found for the most *part in the highest* grade of metamorphism (granulite facies) An aluminous diopside *called omphacite* is a characteristic mineral in the *eclogite facies of* metamorphism *Pyroxenes are* found as gangue *minerals in* certain ore deposits and occur *in* many slags and *meteorites* Acmite and jadeite are associated with sodium rich *rocks In* addition jadeite is thought *to represent the high* pressure transformation of nephaline and albite



$$NaAlSiO_4 + NaAlSi_3O_8 \rightarrow 2NaAl(SiO_3)_2$$

or

$$NaAlSi_3O_8 \rightarrow NaAl(SiO_3)_2 + SiO_2$$

See DIOPSIDE ORTHORHOMBIC PYROXENE.

Crystal structure The compositional relationships of the pyroxenes are best explained from the point of view of their crystal structure The tetrahedral grouping of four oxygen atoms about a central 4-valent silicon atom to form the SiO_4^{4-} anion results in an excess negative charge for the group

This negative charge can be decreased by the sharing of oxygen atoms between neighboring SiO_4 tetrahedra The pyroxenes are characterized by the SiO_4 tetrahedra sharing two oxygen atoms with two neighboring tetrahedra which results in endless polymerized chains of the effective composition SiO_3^{-2}

The remaining charges are satisfied by the 1 2 and 3 valent metal atoms Thus the crystal is formed by the bonding of the silicate chains to parallel silicate chains by these metal atoms Na and Ca occupy equivalent sites in the mineral and can substitute for each other Mg Fe^{+} Fe^{3} and Al occupy equivalent sites and substitute for each other and up to 12% of the Si atoms can be replaced by Al

In nature the pyroxenes often are altered with amphibole chlorite serpentine and talc as the more common alteration products The transformation of pyroxene to hornblende is sometimes extensive and has been referred to as the uralization process the alteration product being called uralite *See* SILICATE PHASE EQUILIBRIA SOLID STATE CHEMISTRY [G W D]

Pyroxenite

A heavy dark colored phaneritic (visibly crystalline) igneous rock composed largely of pyroxene with smaller amounts of olivine or hornblende

Pyroxenite composed largely of orthopyroxene occurs with anorthosite and peridotite in large *banded gabbro bodies These rocks are formed by* crystallization of gabbroic magma (rock melt) Some of these pyroxenite masses are rich sources of chromium Certain pyroxenites composed largely of clinopyroxene are also of magmatic origin but many probably represent products of reaction between magma and limestone Other pyroxene-rich rocks have formed through the processes of meta*morphism and metasomatism See* GABBRO IGNEOUS ROCKS PERIDOTITE PYROXENE [C A CA]

Pyrrhotite

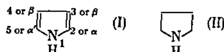
An iron sulfide mineral crystallizing in the hexagonal system Pyrrhotite crystals are rare but when found they are tabular parallel to the base It is *usually massive with a metallic luster and brown*ish bronze color The hardness is 4 (Mohs scale) and the specific gravity 4 6 Pyrrhotite is one of the few minerals that are magnetic that is it will be attracted to the ordinary magnet hence the name magnetic pyrites

The chemical composition of pyrrhotite is written $Fe_{1-x}S$ with x between 0 and 0 2 The mineral troilite is FeS but other varieties have a deficiency of iron that is the ratio of iron to sulfur is less than 1 1 The more iron the less magnetic is the mineral troilite is nonmagnetic

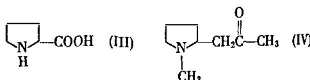
Pyrrhotite is a common minor constituent of many igneous rocks In some basic igneous rocks associated with chalcopyrite and pentlandite it is found in large masses which may have been segregated by magmatic differentiation Pyrrhotite is also found in veins contact metamorphic deposits and pegmatites It is found in large quantities in Finland Norway, Sweden and Russia The most noted occurrence is at Sudbury Ontario Canada, where it is mined on a large scale for the associated pentlandite $(Fe\,Ni)_9S_8$ chalcopyrite $CuFeS_2$ and other nickel and copper minerals *See* PENTLANDITE PYRITE [C S HU]

Pyrrole

One of a group of organic compounds containing a doubly unsaturated five-membered ring in which nitrogen occupies one of the ring positions Pyrrole (I) is a representative compound *See* HETEROCYCLIC COMPOUNDS The pyrrole system is found in



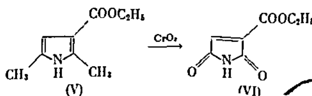
the green leaf pigment chlorophyll in the red blood pigment hemoglobin and in the blue dye indigo (*see* INDOLE PORPHYRIN) Interest in these colored bodies has been largely responsible for *the intensive study of pyrroles Tetrahydropyrrole* or pyrrolidine (II) is part of the structures of two protein amino acids proline (III) and hydroxy



proline and of hygrine (IV) an alkaloid from Peruvian coca

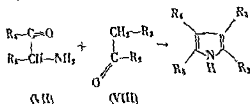
Properties Pyrrole (I) is a liquid bp 130°C n_D^{20} 1 5085 (1 5098) and density (20/4) 0 948 (0 969) that darkens and resinifies on standing in air and that polymerizes quickly when treated with mineral acid Polyalkyl pyrroles are not so sensitive and negatively substituted pyrroles are even less so Pyrrole is a planar aromatic compound with an experimental resonance energy of 22 27 kcal/mole Familiar substitution processes such as halogenation nitration sulfonation and acylation can be realized Substitution generally occurs more readily than in the corresponding benzene analog The en*tering group favors the 2 or 5 position Pyrrole by virtue of its heterocyclic nitrogen is very weakly* basic and is comparable in this respect to urea or to semicarbazide The hydrogen at the 1 position is removable as a proton and accordingly pyrrole is also an acid although a weak one

Pyrroles are not as resistant to oxidation as the analogous benzene compounds Controlled chromic acid oxidation converts pyrroles with or without groups in the α positions to the corresponding maleimides In this way for example ethyl 2 5

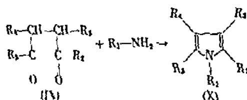


dimethylpyrrole-3-carboxylate (V) is oxidized to a maleimide derivative (VI) Conditions are available for useful reductions of groups attached to the nucleus Zinc and hydrochloric acid can reduce the pyrrole ring to dihydropyrrole (pyrroline) Raney nickel under rigorous conditions, or platinum or palladium under milder conditions, serve as catalysts in the hydrogenation of pyrroles to pyrrolidines

Preparation. The Knorr synthesis probably the most versatile pyrrole synthesis, combines an α aminoketone (VII) with an α methylene carbonyl compound (VIII) The condensation is carried out



either in glacial acetic acid or in aqueous alkali The R groups may be varied widely, however, best results are obtained when R_3 of the carbonyl compound (VIII) is an activating group, for example, carbethoxy Another useful general synthesis the Paal Knorr synthesis converts a 1 4-dicarbonyl compound (IX) by cyclization with ammonia



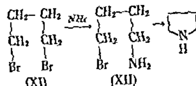
(R_2=H) or with a primary amine to a pyrrole (X) Pyrrole itself is obtained by pyrolysis of ammonium mucate (saccharate) Pyrolysis of primary amine salts of mucic acid gives 1 substituted pyrroles

Derivatives Chloro- bromo- and iodopyrroles have been prepared with bromopyrroles comprising the largest group Halopyrroles carrying an electronegative group such as carbethoxy are more stable than the same compounds without the group 2,3 4 5-Tetraiodopyrrole formed by iodination of pyrrole or of tetrachloromercuripyrrole is useful as an antiseptic in the same areas as iodoform Pyrroles with ethylmagnesium bromide react as active hydrogen compounds to give Grignard derivatives Ethyl formate reacts with such pyrryl Grignard reagents to form pyrrole aldehydes chloroformic ester reacts to give pyrrolecarboxylic esters Pyrrole aldehydes can also be formed by formylation of a vacant pyrrole position with hydrocyanic acid and h[illegible]

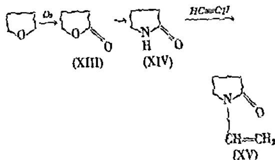
[illegible]

the pyrrole nucleus

Pyrrolidine (III) b p 87-88°C, can be prepared by catalytic hydrogenation of pyrrole or by ring closure reactions Either 1 4-dibromobutane (XI) or 4 bromobutylamine (XII) can serve as starting material 2 Ketopyrrolidine, or pyrrolidone (XIV)



is of considerable interest in connection with the

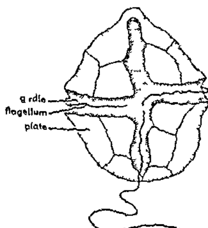


preparation of polyvinylpyrrolidone Pyrrolidone, which can be formed from tetrahydrofuran by autoxidation in the presence of a cobalt catalyst and treatment of the resulting γ butyrolactone (XIII) with ammonia, is combined with acetylene to form vinylpyrrolidone (XV) Polymerization of this material furnishes polyvinylpyrrolidone, a material of relatively high molecular weight, which is suitable for maintaining osmotic pressure in blood and so acting as an extender for plasma or whole blood See PYRIDINE. [W J CE.]

Bibliography R C Elderfield (ed), *Heterocyclic Compounds*, vol 1, 1950

Pyrrophyta

A small phylum of algae characterized by the presence of yellowish green to golden brown plastids, by the storage of food reserves in the form of starch or starchlike compounds, and sometimes oils, and by the general absence of cell walls (but if cell walls are present they usually contain cellulose)



Gonyaulax, a dinoflagellate, with walls consisting of articulated plates Two flagella are attached in transverse groove, or girdle (From H J Fuller and O Tippo, *College Botany*, rev ed, Holt, 1954)

Most members are unicellular and biflagellate A few are without flagella The usual method of reproduction is by cell division although some reproduce by means of zoospores (swimming spores) or nonmotile spores and sexual reproduction has been observed in a few The taxonomy of the Pyrrophyta has been much discussed and various interpretations have been proposed According to the views of several modern algologists the phylum is divided into two classes the Desmophyceae (Desmokontae) and the Dinophyceae

The Desmophyceae consisting of 30 species is a group of rare mostly marine algae somewhat similar to the diatoms (*see* CHRYSOPHYTA, DIATOM) Some of these with or without cell walls are flagellate and motile whereas others have cell walls but no flagella

The Dinophyceae a large group of 1000 species occur in both fresh and salt water and usually as components of plankton The majority of this class are motile unicellular flagellates (dinoflagellates) but there are also some nonflagellated genera (phytodinads) including rare epiphytic and parasitic algae

One of the chief characteristics of the dinoflagellates is the presence of grooves one of which the girdle encircles the cell transversely or spirally and the other of which extends longitudinally along one side only Each of these grooves contains a flagellum In a few genera the protoplasts are naked but most have cellulose walls either homogeneous or made up of a specific number of articulate plates Many of the marine dinoflagellates are phosphorescent and emit enough light to be conspicuous at night when the water is disturbed One of these organisms (*Gymnodinium*) becomes so prevalent at times that it produces what are known as red tides which result in the destruction of large numbers of fish (*see* PALEOECOLOGY) However most of the dinoflagellates together with the diatoms have a basic role in the food economy of organisms living in water [P A V]

Bibliography *See* THALLOPHYTA

Pythagorean theorem

In any right triangle the square on the hypotenuse is equal to the sum of the squares on the other two sides $r^2 = x^2 + y^2$ More than 100 different proofs have been given for this extremely important theorem of euclidean plane geometry which bears the name of the Greek philosopher and mathematician Pythagoras who lived in the sixth century B C One of these is based on a comparison of areas in the squares in Fig 1

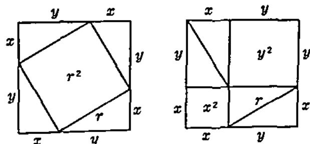


Fig 1 Pythagorean theorem

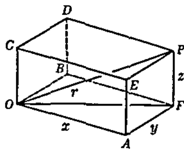


Fig 2 Three-dimensional Pythagorean theorem

The 3-dimensional pythagorean theorem may be phrased the square of the diagonal of a rectangular box is equal to the sum of the squares of three adjacent edges that meet at a vertex $r^2 = x^2 + y^2 + z^2$

Squares that are equal to the sum of two or three squares were known to Pythagoras For example $3^2 + 4^2 = 5^2$ and $1^2 + 4^2 + 8^2 = 9^2$ The name pythagorean triple is given to any triple (x y r) of whole numbers such that $x^2 + y^2 = r^2$ and the name pythagorean quadruple to such integral quadruples (x y z r) for which $x^2 + y^2 + z^2 = r^2$ All pythagorean triples in which x y r are without common factor and x is odd are obtained by replacing the letters a and b in the triple ($a^2 - b^2$ $2ab$ $a^2 + b^2$) by whole numbers that have odd sum and no common factor and similarly all such pythagorean quadruples are found by replacing a b c d in the quadruple ($a^2 - b^2 + c^2 - d^2$ $2ab - 2cd$ $2bc - 2ad$ $a^2 + b^2 + c^2 + d^2$) by whole numbers that have odd sum and no common factor *See* SQUARE TRIANGLE [J S F]

Python

Any of several snakes of the subfamily Pythoninae family Boidae Except for one small species in Mexico they are limited to the warmer portions of the Old World Like the boas they have retained the rudiments of a pelvic girdle and the femurs which are usually visible as a pair of claws near the anus Many pythons are quite large commonly 14–15 ft long and one the reticulated python *Python reticulatus* has been reported to reach lengths up to 32 ft Among the snakes it is exceeded in size only by the heavier anaconda of South America *See* BOA SQUAMATA [J D B]



The python *Python* sp length to 32 ft (*From E L Palmer Fieldbook of Natural History McGraw Hill 1949*)



Q to Quinone

Q

Often called the quality factor of a circuit Q is defined in various ways depending upon the particular application In the simple RL and RC series circuits Q is the ratio of reactance to resistance that is

$$Q = X_L / R \quad \text{or} \quad Q = X_C / R \quad \text{(a numerical value)}$$

where X_L is the inductive reactance X_C is the capacitive reactance and R is the resistance An important application lies in the dissipation factor or loss angle when the constants of a coil or capacitor are measured by means of the alternating current bridge

Q has greater practical significance with respect to the resonant circuit and a basic definition is

$$Q_0 = 2\pi \frac{\text{maximum stored energy per cycle}}{\text{energy lost per cycle}}$$

where Q_0 means evaluation at resonance For certain circuits such as cavity resonators, this is the only meaning Q can have

For the RLC series resonant circuit with resonant frequency f_0

$$Q_0 = 2\pi f_0 L/R = 1/2\pi f_0 CR$$

where R is the total circuit resistance L is the inductance and C is the capacitance Q_0 is the Q of the coil if it contains practically the total resistance R The greater the value of Q_0 the sharper will be the resonance peak

The practical case of a coil of high Q_0 in parallel with a capacitor also leads to $Q_0 = 2\pi f_0 L/R$ R is the total series resistance of the loop although the capacitor branch usually has negligible resistance

In terms of the resonance curve

$$Q = f_0 / (f_2 - f_1)$$

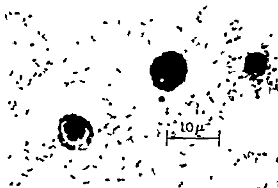
where f_0 is the frequency at resonance and f_1 and f_2 are the frequencies at the half power points See RESONANCE (ALTERNATING CURRENT CIRCUITS)

[B L R]

Q fever

An acute febrile infectious disease in man characterized by sudden onset patchy pneumonitis absence of rash and low mortality and caused by a bacterial like microorganism *Coxiella burnetii* Diagnosis is based on development of serum antibody but the Weil Felix test is negative Persons handling livestock or their products (milk meat wool and fertilizer) are predisposed to infection especially by inhalation while tick transmission to man is rare See RICKETTSIOSES

Coxiella burnetii the causative agent is world wide in distribution Unlike most other rickettsiae the organism has a filterable phase resists desiccation and remains viable for months in soil and in animal and tick wastes A primary cycle between certain rodents such as bandicoots in Australia and their tick parasites occurs in nature Passage

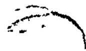


Coxiella burnetii causative agent of Q fever in stained smear of yolk sac from infected chicken embryo (Photomicrograph by N J Kramis)

through the eggs in some species of ticks augments natural maintenance and many species have been found naturally infected Secondary cycles of obscure relationship to the primary cycle occur in domestic animals Infections in these hosts are inapparent and chronic some animals shed organisms in their milk for more than one lactation Placentas of domestic animals are highly infectious and become on disintegration the chief source of infection for man and animals Vaccination of valuable dairy herds gives some promise of reduction of incidence in man and domestic animals [C B P]

Q meter

A direct reading instrument widely used for measuring the Q of an electric circuit at radio frequencies (see Q) Originally designed to measure the Q of coils the Q meter has been developed into a flexible, general purpose instrument for determining many other quantities such as (1) the distributed capacity effective inductance and self resonant frequency of coils (2) the capacitance Q or power factor and self resonant frequency of ca



pacitors, (3) the effective resistance, inductance or capacitance, and the Q of resistors, (4) characteristics of intermediate and radio frequency transformers, and (5) the dielectric constant, dissipation factor, and power factor of insulating materials

The illustration shows in simplified form the measurement circuit of a Q meter The coil L_x being measured is connected into the circuit by means of the external terminals HI and LO The coil is brought into resonance by tuning the calibrated capacitor C_c A controlled and measured input voltage e is introduced into the circuit by means of an rf oscillator A thermocouple voltmeter measures the input voltage e, and a vacuum tube voltmeter (VTVM) measures the voltage V across the calibrated capacitor With the circuit tuned to resonance,

$$\omega L_x - 1/\omega C_c$$

and

$$\frac{V}{e} = \frac{(R_x^2 + \omega L_x^2)^{1/2}}{[R_x^2 + (\omega L_x - 1/\omega C_c)^2]^{1/2}} - (1 + \omega^2 L_x^2/R_x^2)^{1/2}$$

where R_x is the resistance of the coil Since $Q = \omega L/R$

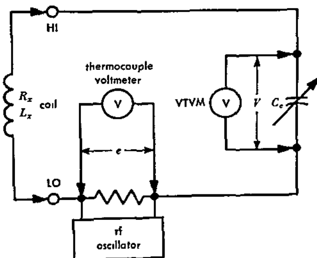
$$\frac{V}{e} = (1 + Q^2)^{1/2}$$

When Q is large the equation may be simplified to this basic equation of the Q meter

$$Q - \frac{V}{e}$$

The thermocouple voltmeter and the vacuum tube voltmeter are calibrated in such a way that the product of their readings gives the Q of the coil directly

The many quantities enumerated above are determined by inserting suitable circuit elements in series with the coil or in parallel with the capacitor



Simplified measurement circuit of a Q meter

and measuring the effect on the circuit Q Simple calculations based on the preceding equations are required to determine the desired quantity from the measured values of Q *See* ELECTRICAL MEASUREMENTS, IMPEDANCE MEASUREMENTS HIGH FREQUENCY [F K]

Bibliography F E Terman and J M Pettit, *Electronic Measurements* 2d ed , 1952

Quadrature

The condition in which the phase angle between two alternating quantities is 90°, corresponding to one quarter of an electrical cycle The electric and magnetic fields of electromagnetic radiation are in space quadrature, which means that they are at right angles in space *See* ELECTROMAGNETIC RADIATION

The current and voltage of a perfect coil are in quadrature because the coil current lags behind the coil voltage by exactly 90° The current and voltage of a perfect capacitor are also in quadrature, but here the current leads the voltage by 90° In these last two cases the current and voltage are in time quadrature *See* ALTERNATING CURRENT CIRCUIT THEORY [J MR]

Quadric surface

A surface defined analytically by an equation of the second degree in three variables If these variables are x, y, z such an equation has the form

$$ax^2 + by^2 + cz^2 + 2exy + 2fxz + 2gyz + 2px + 2qy + 2rz + d = 0 \quad (1)$$

Every plane section of such a surface is a conic To simplify the notation the corresponding capital letter is used to denote the cofactors of the coefficients $a, b, c, d, e, f, g, p, q, r$ in the determinant

$$\Delta = \begin{vmatrix} a & e & f & p \\ e & b & g & q \\ f & g & c & r \\ p & q & r & d \end{vmatrix}$$

Thus $D = abc + 2efg - ag^2 - bf^2 - ce^2$ By λ_1 λ_2 λ_3 are denoted the three roots of the polynomial

$$\phi(\lambda) = \begin{vmatrix} a - \lambda & e & f \\ e & b - \lambda & g \\ f & g & c - \lambda \end{vmatrix}$$

whose product is D

Quadrics are classified as central if $D \neq 0$ and noncentral if $D = 0$, nondegenerate if $\Delta \neq 0$, and degenerate if $\Delta = 0$ For further subclassifications it can be assumed that the coefficients in Eq (1) are real numbers The roots of $\phi(\lambda)$ will then be real also

Central quadrics A central quadric has a center of symmetry at the point $x = P/D$, $y = Q/D$, $z = R/D$ Its equation is freed of linear terms by the translation of axes $x' = x - P/D$ $y' = y - Q/D, z' = z - R/D$ and becomes

$$ax'^2 + by'^2 + cz'^2 + 2ex'y' + 2fx'z' + 2gy'z' = -\Delta/D \quad (2)$$

By an appropriate rotation of axes this equation can be reduced further to the form

$$\lambda_1 X^2 + \lambda_2 Y^2 + \lambda_3 Z^2 = -\Delta/D \qquad \phi(\lambda_i) = 0 \quad (3)$$

If $\Delta = 0$ the degenerate central quadric Eq (1) is called a quadric cone Whenever the real roots λ_i of $\phi(\lambda) = 0$ are not all of the same sign the quadric cone is real and is a surface formed of lines connecting the points of a real conic to a point outside its plane—called the vertex If the roots λ_i are all of the same sign the quadric cone *is imaginary except for a single real point the vertex*

If $\Delta \neq 0$ the nondegenerate central quadric Eq (1) is a real ellipsoid a *hyperboloid of one* sheet a hyperboloid of two sheets or an imaginary ellipsoid according to whether three two one or none of the λ have the same sign as $-\Delta/D$ If two of the λ_i are equal the central quadric is a spheroid or a hyperboloid of revolution If the three *roots λ are equal it is a sphere*

Noncentral quadrics If $D = 0$ the quadric Eq (1) is noncentral and $\phi(\lambda) = 0$ has a root $\lambda_3 = 0$ A suitable rotation of axes reduces the equation to

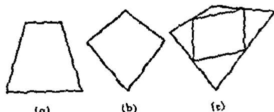
$$\lambda_1 X^2 + \lambda_2 Y^2 + 2\mu_1 X + 2\mu_2 Y + 2\mu_3 Z + d = 0$$

where $\Delta = -\lambda_1\lambda_2\mu_3^2$ A translation further reduces μ_1 to 0 if $\lambda_1 \neq 0$, μ_2 to 0 if $\lambda_2 \neq 0$ and d to 0 if $\mu_3 \neq 0$ A nondegenerate noncentral quadric ($D = 0$ $\Delta \neq 0$) is an elliptic paraboloid if $\lambda_1\lambda_2 > 0$ or a hyperbolic paraboloid if $\lambda_1\lambda_2 < 0$ A degenerate noncentral *quadric* ($D = \Delta = 0$) is either a quadric cylinder (if at least one of the *cofactors in Δ is not 0*) or a pair of planes a single plane or a line or it has no real locus If $\lambda_1\lambda_2 > 0$ and $\mu_3 = 0$ the surface *may be* a real elliptic cylinder a single line or an imaginary elliptic cylinder depending on the value of d

If $\lambda_1\lambda_2 < 0$ and $\mu_3 = 0$ the surface is a hyperbolic cylinder or a pair of planes If either $\lambda_1\mu_3$ and λ_2 are 0 but not both the *surface is* a parabolic cylinder See CYLINDER ELLIPSOID AND SPHEROID HYPERBOLOID PARABOLOID SPHEROID SURFACE AND SOLID OF REVOLUTION [J S F]

Quadrilateral

In the euclidean sense a quadrilateral is a geometric figure bounded by 4 straight line segments called sides each of which intersects each of 2 adjacent sides in points called vertices but fails to intersect the opposite side (considered as a finite segment) The 4 vertices are connected in pairs by the 4 sides and by 2 other segments called diagonals If the latter intersect each other the quadrilateral is a plane quadrilateral if not it is a skew quadrilateral A plane quadrilateral is called a rhombus if its 4 sides are equal or a kite if it has



Quadrilaterals (a) Isosceles trapezoid (b) Kite (c) Parallelogram in quadrilateral

2 pairs of equal adjacent sides A quadrilateral with 1 pair of sides parallel is a trapezoid, if the other pair of sides are equal the trapezoid is called isosceles A quadrilateral with 2 pairs of parallel sides is a parallelogram The midpoints of the sides *of any quadrilateral* plane or skew are vertices of a parallelogram The diagonals of this parallelogram bisect each other in point called the centroid of the quadrilateral *See* PARALLELOGRAM RECTANGLE, SQUARE, TRAPEZOID [J S F]

Quail

Any of a number of American birds also called *bob-white of the family* Phasianidae characterized *by small* compact bodies short tails and terrestrial habits

In the eastern United States there is only one native quail *Colinus virginianus* It is found over most of the United States *except the Far West and* Northwest The bob-white is a brown streaked bird characteristic of the forest border It forms *flocks* of small to moderate size rarely of more than 30 birds usually with about 12 15 individuals *in a covey* This quail is both a valuable game bird and a useful bird to the farmer feeding primarily upon weed seeds and injurious insects



The California quail *Lophortyx californica* length to 9½ in (From E L Palmer *Fieldbook of Natural History* McGraw-Hill 1949)

The western quails are different in many respects Several are characterized by a plume on the head and by more striking plumage generally, often showing black and blue colors *See* GALLIFORMES [J D B]

Qualification test

A formally defined series of tests by which the functional, environmental, and reliability performance of a component or system may be evaluated in order to satisfy the contractor as to its satisfactory design and construction prior to his approval and acceptance In all but the simplest of systems it is impossible to anticipate the wide variety of circumstances under which the component and the system will be expected to perform satisfactorily Therefore, it is essential that the contractor and the contractee for a component or system have a mutual understanding as to how the component or system must perform under specified circumstances in order to be satisfactory *See* SYSTEMS ENGINEERING

Specific criteria are prescribed for functional and environmental characteristics and reliability criteria are established (*see* ENVIRONMENTAL TEST, RELIABILITY OF EQUIPMENT) The tests are usually carried out under formal conditions in the presence of witnesses from the manufacturer and the customer, especially in the case of military equipment Careful formal records are kept of all performance data Formal acceptance of the component or system by the customer is contingent upon the successful completion of the qualification tests [R W M]

Qualitative chemical analysis

The branch of analytical chemistry concerned with finding the nature of all of the substances in a given material (complete analysis) or with the detection of the presence or absence of certain substances of interest

A distinction can be made between the preliminary investigation, consisting of inspection and some simple tests and the systematic chemical procedure of analysis The preliminary investigation together with a careful review of the history of the material under investigation may greatly simplify the analysis For the systematic analysis a part (sample) of the material to be investigated (unknown) is subjected to the necessary preliminary treatment, then the identification or confirmatory tests are applied to indicate whether or not certain substances are present Since it would be too tedious to test individually for the presence or absence of every imaginable substance classification (group) tests or separation procedures are applied first These sometimes permit the omission *of tests for whole classes and groups of substances*

The preliminary treatment *This may be as* simple as grinding and dissolution or it may include fusion and involved procedures of separation *See* CHROMATOGRAPHY, EXTRACTION, ION EXCHANGE, PRECIPITATION (CHEMISTRY), SAMPLING TECHNIQUES, SEPARATION (CHEMICAL AND PHYSICAL), SOLUBILIZING OF SAMPLES, VOLATILIZATION

A report on qualitative analysis must contain some quantitative information, that is, indications concerning the approximate relative amounts of the substances present This may be done by indicating major constituents (majors) representing 5–100% of the material, minor constituents (minors), representing 0 1–5% of the material and trace constituents (traces), each accounting for less than 0 1% of the material A statement that a sample consists mostly of copper and zinc, with a minor amount of tin and traces of iron and sulfur will identify the material as a brass, whereas omission of the quantitative information would leave a choice among impure samples of the five elements the four possible sulfides, and all alloys containing some of the four metals The quantitative information is gained by starting with known amounts of sample, estimating the volumes of separating phases, evaluating the intensities of color reactions, and using sensitivity data

Quantitative information is needed also because one cannot assume the absolute absence of all the substances that are not reported present Analytical testing cannot assure absence beyond the limitations of the most sensitive test applied The reporting of traces of common substances must be supported by the most rigorous performance of blanks (tests without the unknown) What constitutes a common material is determined to some extent by the routine work of a laboratory, its location and construction, and the apparatus and chemicals used Furthermore negative findings concerning expected minors or traces should be proved by controls, that is, by analyses of samples to which the expected substance has been added

Selectivity and sensitivity. Most chemical tests result in changes affecting the transmission of light by the appearance or disappearance of phases (solids liquids or gases) and their boundaries that is changes from homogeneity to heterogeneity and vice versa, as in precipitation, dissolution evolution of gas condensation or evaporation, or in distinct changes in color which affect the absorption of light

The practical value of a test depends mainly upon its selectivity which is determined by the number of substances responding in a like manner A test given by only a small number of substances is called selective Tests procedures or reagents to which only one substance responds in a certain way are called specific

Sensitivity is the general term for the ability of a test, procedure or reagent to show the presence *of small amounts of a substance It is described by* stating the *limiting dilution the limit of identification* and the limiting proportions

The limiting dilution D is defined as the lowest concentration of the substance that always gives a positive test In place of the limiting dilution the

negative logarithm pD is sometimes used These are simpler figures that increase with sensitivity The term pDa indicates the limiting dilution when no interfering substances are present

The limit of identification (LI) is the smallest absolute quantity (mass or volume) that always gives a positive test It is approximately related to the limiting dilution since the LI becomes smaller as the scale of a test diminishes The limiting dilution corresponding to pDa = 3.5 is 0.0003 g/ml, and the LI becomes 0.3 mg if 1 ml of solution is taken for the test If the test is performed with 1 drop of 0.03 ml volume the LI decreases to 0.009 mg Finally if means are found to perform the test with 0.0003 ml = 0.3 μl of solution the LI becomes 0.0001 mg = 0.1 μg

The *limiting proportions* (LP) are the smallest ratios (mass of substance sought/mass of interfering substance present) in which the sought substance *is still detectable* provided that there is enough *of the sought substance* to equal or exceed the limiting dilution and the limit of identification. Both *of the latter may* be affected by the presence of interfering substances and it has been suggested to list experimentally determined dilution exponents pDr with the limiting proportions Thus for the sodium test with uranyl acetate pDa = 5.0 1 90 NH_4 with pDr = 4.2 This means that the *sodium test* is still obtained if the quantity of ammonium ion present is 90 times that of the sodium ion but that the concentration sensitivity of the sodium test suffers somewhat

Any companion (cosolute) will interfere if both it and the unknown give like or similar effects with the reagent Some companions use up the reagent without producing any noticeable effect in such instances more reagent must be added so that some is left to react with the substance sought Other companions unite with the sought substance and thus hinder its action Whether or not a plausible explanation for the interference is available the limiting proportions are of great practical importance since they indicate how much of the unknown can be detected in the presence of given amounts of given cosolutes

Systematic qualitative analysis This step is frequently based *upon a combination of two* possible procedures a system of classification and identification tests on a large *number of* separate samples and systematic separation and *isolation from* one sample of the substances the nature of which is finally confirmed by identification tests

Identification by tests with individual samples of the unknown without any *preliminary* separations must start with classification and *group tests* These permit the elimination of a large number of *possibilities* This still may leave a *considerable* number of individual *identification tests for all* the substances *belonging to groups* that have not been excluded The procedure has the disadvantage *that* many samples of the unknown will be needed *if the latter is a* complex mixture

Systematic separation is based upon the precipitation distillation or extraction of some of the substances *from the* sample whereas *other* substances remain in the original solution Groups of substances are first isolated then split into subgroups, and finally into fractions of subgroups each of which contains only one or a *few* substances that may then be identified by confirmatory tests The advantages of the procedure are that *one sample* suffices for the detection of most of the substances and that these *substances are obtained* in isolated form available for additional testing if necessary

The sensitivity of detection is limited for each substance by that *step in the whole procedure* (classification or group test, method of separation or isolation identification or confirmatory *test*) which has the *poorest limiting* proportion limiting dilution or limit of *identification* For example if a sample contains less than one weight *of calcium* oxide for one hundred weights *of magnesium* oxide a precipitate *of calcium oxalate* fails to separate on adding ammonium *oxalate to the alkaline* test solution This limiting *proportion* of the calcium oxalate *test* (1 100 Mg) will prevent the detection of an amount of calcium *less than one* one-hundredth of *the magnesium* present whether the reaction is used to test for the alkaline earth ions (Ca^{++}, Sr^{++}, Ba^{++}) or to separate them from the magnesium *This will be true* no matter what the sensitivity of the *confirmatory* test for calcium

Since information on limiting *proportions* of *separations and tests* is still quite incomplete, the first qualitative analysis of a very complex mixture of entirely unknown *character cannot* discover more than the majors, all or most of the minors *and possibly a few* trace constituents A repetition will be required using separations and tests with more *favorable* limiting proportions to make certain of all of the *minors* and *to detect more traces* The search for small traces of special interest *the final* step would have to *be carried* out with *methods* specially adapted to the material *on* hand *See*

lyzed in several ways *Blow pipe analysis* an early technique whereby *an inorganic* sample was added to a bead of low melting *material* such as borax *and the resulting color of the bead* noted after removal of the bead *from a flame* has been largely discarded *See* CHEMICAL MICROSCOPY, FLAME PHOTOMETRY, MAGNETIC RESONANCE, SPECTRO CHEMICAL ANALYSIS SPECTROPHOTOMETRIC ANALYSIS SPOT TEST ANALYSIS X RAY DIFFRACTION, X RAY FLUORESCENCE ANALYSIS

In the investigation of organic materials substances *may be identified by their physical* properties such as refractive index melting point, transition temperatures, and distribution behavior between two phases *See* ORGANIC QUANTITATIVE ANALYSIS

The western quails are different in many respects Several are characterized by a plume on the head and by more striking plumage generally often showing black and blue colors *See* GALLIFORMES
[J D B]

Qualification test

A formally defined series of tests by which the *functional environmental* and *reliability* performance of a component or system may be evaluated in order to satisfy the contractor as to its satisfactory design and construction prior to his approval and acceptance In all but the simplest of systems it is impossible to anticipate the wide variety of circumstances under which the component and the system will be expected to perform satisfactorily Therefore it is essential that the contractor and the contractee for a component or system have a mutual understanding as to how the component or system must perform under specified circumstances in order to be satisfactory *See* SYSTEMS ENGINEERING

Specific criteria are prescribed for functional and environmental characteristics and reliability criteria are established (*see* ENVIRONMENTAL TEST RELIABILITY OF EQUIPMENT) The tests are usually carried out under formal conditions in the presence of witnesses from the manufacturer and the customer especially in the case of military equipment Careful formal records are kept of all performance data Formal acceptance of the component or system by the customer is contingent upon the successful completion of the qualification tests
[R W M]

Qualitative chemical analysis

The branch of analytical chemistry concerned with *finding the nature of all of the substances in a given* material (complete analysis) or with the detection of the presence or absence of certain substances of interest

A distinction can be made between the preliminary investigation consisting of inspection and some simple tests and the systematic chemical procedure of analysis The preliminary investigation together with a careful review of the history of the material under investigation may greatly simplify the analysis For the systematic analysis a part (sample) of the material to be investigated (unknown) is subjected to the necessary preliminary treatment then the identification or confirmatory tests are applied to indicate whether or not certain substances are present Since it would be too tedious to test individually for the presence or absence of every imaginable substance classification (group) tests or separation procedures are applied first These sometimes permit the omission of tests for whole classes and groups of substances

The preliminary treatment This may be as simple as grinding and dissolution or it may include fusion and involved procedures of separation *See* CHROMATOGRAPHY EXTRACTION ION EXCHANGE PRECIPITATION (CHEMISTRY) SAMPLING TECHNIQUES SEPARATION (CHEMICAL AND PHYSICAL) SOLUBILIZING OF SAMPLES VOLATILIZATION

A report on qualitative analysis must contain some quantitative information that is indications concerning the approximate relative amounts of the substances present This may be done by indicating major constituents (majors) representing 5–100% of the material minor constituents (minors) representing 0 1 5% of the material and trace constituents (traces) each accounting for less than 0 1% of the material A statement that a sample consists mostly of copper and zinc with a minor amount of tin and traces of iron and sulfur will identify the material as a brass whereas omission of the quantitative information would leave a choice among impure samples of the five elements the four possible sulfides and all alloys containing some of the four metals The quantitative information is gained by starting with known amounts of sample estimating the volumes of separating phases evaluating the intensities of color reactions and using sensitivity data

Quantitative information is needed also because one cannot assume the absolute absence of all the substances that are not reported present Analytical testing cannot assure absence beyond the limitations of the most sensitive test applied The reporting of traces of common substances must be supported by the most rigorous performance of blanks (tests without the unknown) What constitutes a common material is determined to some extent by the routine work of a laboratory its location and construction and the apparatus and chemical used Furthermore negative findings concerning expected minors or traces should be proved by controls that is, by analyses of samples to which the expected substance has been added

Selectivity and sensitivity Most chemical tests result in changes affecting the transmission of light by the appearance or disappearance of phases (solids liquids or gases) and their boundaries that is changes from homogeneity to heterogeneity and vice versa as in precipitation dissolution evolution of gas condensation or evaporation or in distinct changes in color which affect the absorption of light

The practical value of a test depends mainly upon its selectivity which is determined by the number of substances responding in a like manner A test given by only a small number of substances is called selective Tests procedures or reagents to which only one substance responds in a certain way are called specific

Sensitivity is the general term for the ability of a test procedure or reagent to show the presence of small amounts of a substance It is described by stating the limiting dilution the limit of identification and the limiting proportions

The limiting dilution D is defined as the lowest concentration of the substance that always gives a positive test In place of the limiting dilution the

negative logarithm pD is sometimes used These are simpler figures that increase with sensitivity The term pDa indicates the limiting dilution when no interfering substances are present.

The limit of identification (LI) is the smallest absolute quantity (mass or volume) that always gives a positive test. It is approximately related to the limiting dilution since the LI becomes smaller as the scale of a test diminishes The limiting dilution corresponding to pDa = 3.5 is 0.0003 g/ml and the LI becomes 0.3 mg if 1 ml of solution is taken for the test. If the test is performed with 1 drop of 0.03 ml volume the LI decreases to 0.009 mg. Finally if means are found to perform the test with 0.0003 ml = 0.3 μl of solution, the LI becomes 0.0001 mg = 0.1 μg

The limiting proportions (LP) are the smallest ratios (mass of substance sought/mass of interfering substance present) in which the sought substance is still detectable provided that there is enough of the sought substance to equal or exceed the limiting dilution and the limit of identification. *Both of the* latter may be affected by the presence of interfering substances and it has been suggested to li t experimentally determined dilution exponents pDr with the limiting proportions Thus, for the *sodium test* with uranyl acetate pDa = 5.0 1 90 NH_4 with pDr = 4.2. This means that the sodium test is still obtained if the quantity *of ammonium* ion present is 90 times that of the sodium ion but that the concentration sensitivity of the sodium test suffers somewhat.

Any companion (cosolute) will interfere if both it and the unknown give like *or similar* effects with the reagent. Some companions use up the reagent without producing any noticeable *effect* in such instances more reagent must be added so that some is left to react with the substance sought. Other companions unite with the sought substance and thus hinder its action Whether or not a plausible explanation *for the* interference is available the limiting proportions are of great practical importance since they indicate how much of the unknown *can* be detected in the presence of given amounts of given cosolutes

Systematic qualitative analysis. This step is

... the substances, the nature of which is finally confirmed by identification tests

Identification by tests with individual samples of the unknown without any preliminary separations must start with classification *and group tests* These permit the elimination of a large number of possibilities This still may leave a considerable number of individual identification tests for all the substances belonging to groups that have not been excluded. The procedure has the disadvantage that many samples of the unknown will be needed if the latter is a complex mixture.

Systematic separation is based upon the precipitation distillation or extraction of some of the substances from the sample whereas other substances remain in the original solution Groups of substances are first isolated then split into subgroups and finally into fractions of subgroups, each of which contains only one or a few substances that may then be identified by confirmatory tests. The advantages of the procedure are that one sample suffices for the detection of most of the substances and that these substances are obtained in isolated form available for additional testing if necessary

The sensitivity of detection is limited *for each* substance by that step in the whole *procedure* (classification or group test, *method of separation* or isolation, identification or confirmatory test) which has the poorest limiting proportion limiting dilution or limit of identification For example, if a sample contains less than *one weight of calcium* oxide for one hundred weights of magnesium oxide, a precipitate of calcium *oxalate fails* to separate on adding ammonium oxalate to the alkaline test solution This limiting *proportion of the* calcium oxalate test *(1 100 Mg)* will prevent the detection of an amount of calcium less than *one one-hundredth* of the magnesium present whether the reaction is used to test for the *alkaline* earth ions (Ca^{++}, Sr^{++}, Ba^{++}) *or to separate them from the magnesium* This will be true *no matter* what the sensitivity of the *confirmatory* test for calcium

Since *information* on limiting proportions of separations and tests is still quite incomplete, the first qualitative analysis *of a* very complex mixture of *entirely unknown* character cannot discover *more* than the majors, all *or* most of the minors and possibly a *few trace* constituents A repetition will be required using separations and tests with more favorable limiting proportions to make certain of all *of the* minors and *to detect* more traces The search for small traces of special interest, the final step would have to be carried out with methods specially adapted to the material on hand *See* TRACE ANALYSIS

Methods and techniques Depending upon the nature of the task, inorganic materials may be analyzed in several ways Blow pipe analysis an early technique whereby an inorganic sample was added to a bead of low *melting* material such as borax and the resulting color of *the bead* noted after removal of the bead from a flame, has been largely discarded. *See* CHEMICAL MICROSCOPY, FLAME PHOTOMETRY MAGNETIC RESONANCE, SPECTROCHEMICAL ANALYSIS SPECTROPHOTOMETRIC ANALYSIS SPOT TEST ANALYSIS X RAY DIFFRACTION, X RAY FLUORESCENCE ANALYSIS

In the investigation of *organic* materials substances may be identified by their physical *proper*ties such as refractive index *melting* point, transition temperatures and distribution behavior between two phases *See* ORGANIC QUANTITATIVE ANALYSIS

The principal domain of qualitative chemical analysis is the investigation of small samples, 0 1 mg to 0 1 μg in mass, which may be dust particles, inclusions in structural materials, contaminations or corrosion products on small parts or some unique specimen of particular interest Tasks of this kind call for utmost economy with the unknown, and they do not permit arbitrary limitation to common substances or to inorganic or organic materials Nondestructive physical methods may not be practical with a small specimen or may not furnish sufficient information On the other hand, the chemical approach is quite useful since very small specimens, as a rule, may be expected to consist of just one material, so that a cleverly devised series of separations will in practice reduce the task to a mere system of classification and identification

A thoroughly realistic approach based upon careful study of the history of the specimen and continuous reference to the literature is best If the specimen seems to be a crystal or crystal fragment, a good procedure would be to determine the optical properties and to study the solubility behavior by exposing a sample to the vapors of solvents Soft crystalline matter is easily identified if its melting point occurs below 300°C for example organic substances and many inorganic hydrates This test may also reveal other transition points, and distillation, sublimation or decomposition at low temperature

For further classification of an apparently inorganic material a sample may be heated up to 800°C in narrow glass tubing while first air, then hydrogen chlorine, and finally hydrogen sulfide are passed through the tubing Observation of the changes on the specimen and investigation of the evolved condensates, sublimates and gases may provide definite clues which may be confirmed by simple chemical tests on the residue or condensate left in the tubing Organic matter will be recognized on heating in air and may not be too difficult to identify after consideration of optical properties and transition temperatures that are already known

If the specimen should turn out to have a complex composition (glass alloy solid solution), the problem that led to the analysis will usually be solved by the disclosure of majors and minors which will be found with suitable procedures Methods for the separation of both common and rare elements have been developed and tested on a small scale Any desirable qualitative or quantitative method of separation may be applied to very small samples since the chemical behavior is not noticeably affected by the absolute quantities involved *See* ANALYTICAL CHEMISTRY [A A B P]

Bibliography N D Cheronis and J B Entrikin, *Semimicro Qualitative Organic Analysis,* 2d ed, 1957, W C McCrone, *Fusion Methods in Chemical Microscopy,* 1957, T Moeller, *Qualitative Analysis,* 1958

Quality control

A system of inspection, analysis and action applied to a manufacturing operation so that, by inspecting a small portion of the product currently produced, an estimate of the over all quality of the product can be made to determine what, if any, changes must be made in the operation to achieve or maintain the required level of quality An inspector selects at random a small portion of the output made under production conditions with regular production tools He inspects the selected portion in the conventional manner, using the gages and tools regularly used for this purpose

Inspection. Critical examination of a product to determine its conformance to applicable quality standards or specifications is inspection As a result of inspection, products are accepted or rejected depending on their degree of conformance to applicable quality standards Because inspection is a post mortem operation it has no effect on the quality of products, the inspector accepts or rejects products of the same quality as submitted for inspection

An example of inspection may be observed in the receiving department Here the inspector is presented products received from vendors and he is required to make a decision to accept or reject them as a result of his inspection Regardless of his decision, the quality of the products is not altered

Control. If the results of inspection are communicated to others for action quality can be controlled, as when the receiving department notifies a vendor through the purchasing office The average quality and the range of variation in quality of the sample are compared with applicable standards to determine the degree of conformance If the average quality is not satisfactory it is an indication that the process or tools require adjustments If the range of variation in quality is not satisfactory, it is an indication that the process or tools in their present condition are not capable of producing the desired quality In either case the foreman or line supervisor has the responsibility of effecting changes in the operation to bring about the desired improvement Quality control functions to prevent the production of products outside the applicable quality standards For quality control to function effectively there must be close communication and cooperation between the quality control department and the manufacturing organization

Requirements for control Certain prerequisites must be met before quality control can function effectively The degree by which these prerequisites are met determines the success or failure of the quality control program Principal among these prerequisites are the following

Management policy that quality shall be controlled is important Quantity without quality is not effective, hence, when quality and quantity are in conflict, wise management allows quality to prevail

Quality standards must be clearly and understandably defined The manufacturing organization the quality control department and the customer must have the same concept of the quality standard

Sufficient inspection coverage must be provided There must be enough skilled inspection personnel to inspect the product analyze inspection results and motivate the manufacturing organization to take whatever corrective action is indicated

Proper inspection methods must be used Generally inspection by variables as when diameter of a shaft is measured by a micrometer is more efficient than inspection by attributes as when shaft diameter is measured by a go-not go gage for purposes of quality control

Inspection tools suitable for the job at hand must be available to the inspector A metrological standards laboratory is a necessary adjunct of the quality control department

Adequate records must be kept to show the present status of quality the trend of quality and the costs associated with the quality control effort

Benefits from control The quality control burden may be expected to amount to 10-20% of the direct labor cost However direct benefits may be expected from the effective operation of a quality control department among which are the following

With quality control production increases of 10-35% may be expected from the same facilities over what was accomplished without quality control

Lower manufacturing costs result from elimination of qualifying operations use of equipment and personnel to the optimum capacity and reduction in scrap

Improved employee morale results when management provides the climate for the production of optimum quality there is less criticism of operating or supervisory personnel because of excessive numbers of rejections

Quality is improved with a resulting increase in customer acceptance and a decrease in field complaints and returns

Organization for control The quality control department should be organized along accepted management principles Experience shows that quality control operates most effectively when its manager reports in a staff capacity to the top executive responsible for manufacturing generally the works manager The works manager is responsible for both quantity and quality of production A typical organization chart for a plant emplo... 2000-...

The ... rectly ... uct c... ...ed as an endless chain (Fig 2) The quality control department s responsibility does not end with the shipment of the product It follows product performance in the field and by making specific recommendations motivates the engineering department to redesign the product to improve its quality and performance

Process control methods Both statistical techniques and procedures not requiring the use of statistics are designed to control the operation so that no product will be outside the applicable quality standards

In the receiving department the quality control department inspects all raw materials and purchased finished parts and components A record is kept of the results of these inspections. Periodically the records are reviewed and a process average is calculated for each supplier This process average may be based on the ratio of rejected lots to total lots received or on number of units rejected to number of units received This information is furnished to the purchasing department for discussion with the supplier who if necessary takes steps to improve quality

The feeder departments or parts manufacturing departments, furnish the next opportunity for process control This includes foundries die-casting plastic molding punch press screw machine and other departments which make parts for the

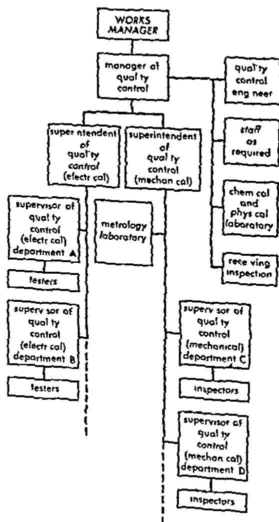


Fig 1 Typical organization chart of quality control department

The principal domain of qualitative chemical analysis is the investigation of small samples, 0 1 mg to 0 1 μg in mass, which may be dust particles, inclusions in structural materials contaminations or corrosion products on small parts, or some unique specimen of particular interest Tasks of this kind call for utmost economy with the unknown, and they do not permit arbitrary limitation to common substances or to inorganic or organic materials Nondestructive physical methods may not be practical with a small specimen or may not furnish sufficient information On the other hand, the chemical approach is quite useful since very small specimens, as a rule, may be expected to consist of just one material so that a cleverly devised series of separations will in practice reduce the task to a mere system of classification and identification

A thoroughly realistic approach based upon careful study of the history of the specimen and continuous reference to the literature is best If the specimen seems to be a crystal or crystal fragment, a good procedure would be to determine the optical properties and to study the solubility behavior by exposing a sample to the vapors of solvents Soft crystalline matter is easily identified if its melting point occurs below 300°C, for example, organic substances and many inorganic hydrates This test may also reveal other transition points and distillation, sublimation or decomposition at low temperature

For further classification of an apparently inorganic material, a sample may be heated up to 800°C in narrow glass tubing while first air, then hydrogen chlorine and finally hydrogen sulfide are passed through the tubing Observation of the changes on the specimen and investigation of the evolved condensates sublimates and gases may provide definite clues which may be confirmed by simple chemical tests on the residue or condensate left in the tubing Organic matter will be recognized on heating in air and may not be too difficult to identify after consideration of optical properties and transition temperatures that are already known

If the specimen should turn out to have a complex composition (glass alloy, solid solution) the problem that led to the analysis will usually be solved by the disclosure of majors and minors which will be found with suitable procedures Methods for the separation of both common and rare elements have been developed and tested on a small scale Any desirable qualitative or quantitative method of separation may be applied to very small samples since the chemical behavior is not noticeably affected by the absolute quantities involved See ANALYTICAL CHEMISTRY [A A B P]

Bibliography N D Cheronis and J B Entrikin, *Semimicro Qualitative Organic Analysis*, 2d ed, 1957, W C McCrone, *Fusion Methods in Chemical Microscopy*, 1957, T Moeller, *Qualitative Analysis*, 1958

Quality control

A system of inspection, analysis and action applied to a manufacturing operation so that, by inspecting a small portion of the product currently produced an estimate of the over all quality of the product can be made to determine what, if any, changes must be made in the operation to achieve or maintain the required level of quality An inspector selects at random a small portion of the output made under production conditions with regular production tools He inspects the selected portion in the conventional manner, using the gages and tools regularly used for this purpose

Inspection. Critical examination of a product to determine its conformance to applicable quality standards or specifications is inspection As a result of inspection, products are accepted or rejected depending on their degree of conformance to applicable quality standards Because inspection is a post mortem operation, it has no effect on the quality of products, the inspector accepts or rejects products of the same quality as submitted for inspection

An example of inspection may be observed in the receiving department Here the inspector is presented products received from vendors and he is required to make a decision to accept or reject them as a result of his inspection Regardless of his decision, the quality of the products is not altered

Control If the results of inspection are communicated to others for action, quality can be controlled, as when the receiving department notifies a vendor through the purchasing office The average quality and the range of variation in quality of the sample are compared with applicable standards to determine the degree of conformance If the average quality is not satisfactory, it is an indication that the process or tools require adjustments If the range of variation in quality is not satisfactory, it is an indication that the process or tools in their present condition are not capable of producing the desired quality In either case the foreman or line supervisor, has the responsibility of effecting changes in the operation to bring about the desired improvement Quality control functions to prevent the production of products outside the applicable quality standards For quality control to function effectively, there must be close communication and cooperation between the quality control department and the manufacturing organization

Requirements for control. Certain prerequisites must be met before quality control can function effectively The degree by which these prerequisites are met determines the success or failure of the quality control program Principal among these prerequisites are the following

Management policy that *quality shall be controlled* is important Quantity without quality is not effective, hence, when quality and quantity are in conflict, wise management allows quality to prevail

does not exceed the corresponding tabulated acceptance number

Reduced inspection is employed when 10 consecutive lots have been accepted under normal inspection Tightened inspection is used when 2 or more lots in 10 have been rejected under normal inspection When 10 consecutive lots have been accepted under tightened inspection, or when a lot is rejected under reduced inspection normal inspection is resumed

Inspection by variables A typical variables inspection plan is illustrated in Figs 3 and 4 The method of recording inspection data and making the necessary calculations is shown in Fig 3 The point determined by the abscissa $A\bar{\bar{d}} - 1$ and the ordinate $0\,43A\bar{R}$ is plotted as shown in Fig 4 If the point falls under the curve the lot is accepted, if the point falls above the curve the lot is rejected

The t test of significance It is often necessary to compare two sample means $\bar{X}_1$ and $\bar{X}_2$ obtained from a pair of random samples of n_1 and n_2 observations with a population in which the standard deviation is σ From these samples can be computed an estimate S of σ by the formula

$$S = \sqrt{\frac{\Sigma_1(X - \bar{X}_1)^2 + \Sigma_2(X - \bar{X}_2)^2}{n_1 + n_2 - 2}}$$

where Σ_1 means to sum the deviations of all observations in the first sample Σ_2 means to sum the deviations of all observations that have been made in the second sample

To determine whether the observed difference $(\bar{X}_1 - \bar{X}_2)$ could have occurred by chance as a result of the inherent variation in the population compute

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

If t is no larger than the t found in the list for $n_1 + n_2 - 2$ degrees of freedom the difference between $\bar{X}$ and $\bar{X}$ [illegible]

95% because a significant t (one which is greater than the tabled value) occurs by chance alone only 1 time in 20 A list of t values for indicated degrees of freedom for a confidence level of 95% follows

Table 1 Single sampling plans by attributes*

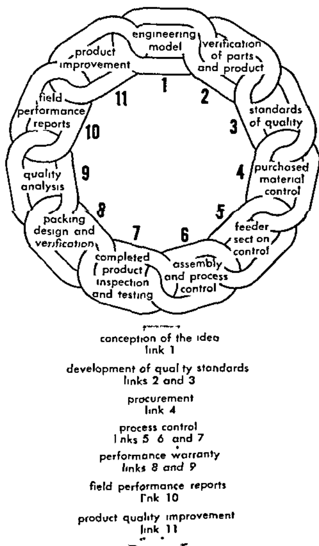
Lot size	AQL											
	65		1 0		1 5		2 5		4.0		6 5	
	n	*Ac*	*n*	*Ac*	*n*	*Ac*	*n*	*Ac*	*n*	*Ac*	*n*	*Ac*
Tightened inspection												
2 8	All	0	All	0	All	0	All	0	7	0	5	0
9 15	All	0	All	0	All	0	10	0	7	0	5	0
16 25	All	0	All	0	15	0	10	0	7	0	5	0
26-40	All	0	25	0	15	0	10	0	7	0	5	0
41-65	35	0	25	0	15	0	10	0	7	0	15	1
66-110	35	0	25	0	15	0	10	0	25	1	15	1
111 180	35	0	25	0	15	0	35	1	25	1	25	2
181 300	35	0	25	0	50	1	35	1	35	2	35	3
301 500	35	0	75	1	50	1	50	2	50	3	50	4
Normal inspection												
2-8	All	0	All	0	All	0	7	0	5	0	3	0
9 15	All	0	All	0	All	0	7	0	5	0	3	0
16-25	All	0	15	0	10	0	7	0	5	0	3	0
26-40	25	0	15	0	10	0	7	0	5	0	10	1
41-65	25	0	15	0	10	0	7	0	15	1	10	1
66-110	25	0	15	0	10	0	25	1	15	1	15	2
111 180	25	0	15	0	35	1	25	1	25	2	25	3
181 300	25	0	50	1	35	1	35	2	35	3	35	5
301 500	75	1	50	1	50	2	50	3	50	4	50	6
Reduced inspection												
2-8	5	0	3	0	2	0	5	1	3	1	2	1
9 15	5	0	3	0	2	0	5	1	3	1	2	1
16-25	5	0	3	0	2	0	5	1	3	1	2	1
26-40	5	0	3	0	2	0	5	1	3	1	2	1
41-65	5	0	3	0	2	0	5	1	3	1	2	1
66-110	5	0	3	0	2	0	5	1	3	1	3	1
111 180	5	0	3	0	7	1	5	1	5	1	5	2
181 300	5	0	10	1	7	1	7	1	7	2	7	3
301 500	15	1	10	1	10	1	10	2	10	2	10	3

* Based on Mil Std 105A Inspection Level No 2 n is sample size Ac is acceptance number

assembly departments Any of several process control procedures may be used, depending on circumstances

First piece inspection is made after the operator has made what he considers a satisfactory setup and has produced what he considers satisfactory products This is essentially a check that the operator understands his job, that he has the correct tools, and that the combination of tools and operator will produce the required quality The responsibility for quality rests essentially on the operator, the inspector provides added assurance against an excessive number of defective units

The $\overline{X}$ and R chart is used in the control of measurable characteristics such as dimensions, weight, density, hardness, or tensile strength Samples of four or five units are taken from the process at regular intervals The average $\overline{Y}$ and the range R of variability of each sample are calculated and plotted on a chart (*see* CONTROL CHART) Any average value or range value falling outside the control limits must be explained and proper corrective action must be taken to eliminate the cause



conception of the idea
link 1

development of quality standards
links 2 and 3

procurement
link 4

process control
links 5 6 and 7

performance warranty
links 8 and 9

field performance reports
link 10

product quality improvement
link 11

Fig 2 Quality control chain showing functions of quality control department

VARIABLES INSPECTION DATA CARD

01-435 | 12 | 2 | 335664 | 423-07112 | 3 | 300
Insert | Length
S # 887632 Base | Dept 31 | 2 130 | 2 120
Greenville Machine Co

076 / 20 = 0038

017 / 4 = 00425

1 / (2 130 − 2,120) = 1 / 0010 = 200

200 × 0038 = −0.24

200 × 00425 = 365

X | D L Q 1 29 57

Fig 3 Manner of recording data and making calculations for variables sampling plan

The *p* chart is useful when the data are a result of testing every unit of product or when relatively large samples are inspected by methods of attributes When several types of repetitive defects are found on products, the frequency of occurrence of each defect as well as the fraction defective can be recorded on a *p* chart In this manner the most significant defect can be recognized and a suitable project can be initiated to correct the cause

Sampling tables. Detail inspection or 100% inspection, on the average, can pick out about 80% *of the possible defects in units in a submitted lot* Generally, 100% inspection means 100% inspection of some one or two quality characteristics and no inspection of the several other characteristics which may be important

Sampling requires the inspection of a relatively small quantity of units in a submitted lot with the lot's being accepted or rejected on the quality of the sample The sample is picked at random so that it might be truly representative of the lot and it is inspected more completely and with more precision than would be possible for the whole lot Inspection may be by attributes or by variables

All sampling involves some risk (the consumer's risk) of passing lots which contain a higher percentage of defectives than the acceptable quality level (AQL) Conversely, there is also the risk (the producer's) of rejecting some lots containing a smaller percentage of defectives than the prescribed AQL

Inspection by attributes A typical set of sampling plans by attributes is shown in Table 1 A sample of the tabulated normal number of units *n* is inspected, the size sample depends on the size lot from which it is taken and the AQL. The lot is accepted if the number of defectives in the sample

chemical methods The procedures followed in making an analysis may be classed according to (1) the type of constituent determined (2) the kind of method used (3) the type of material analyzed and (4) the amount of constituent present

Classification of procedures In an ultimate analysis the amount of a single element or compound is determined In a proximate analysis certain constituents are determined as a group of indefinite relative composition—for example the determination of ash in a sample of coal

Methods of analysis may be direct or indirect In a direct gravimetric method the desired constituent is converted to a compound of definite known composition and this compound is weighed In a direct volumetric method the desired constituent is determined by measuring the volume of reagent of known concentration required to react completely with the constituent

In an indirect gravimetric method a mixture of substances which includes the desired constituent is weighed and then wholly or in part converted to some other substance or mixture of substances of known composition and weighed The amount of desired constituent can then be calculated by solving two simultaneous equations that can be set up from the data obtained Indirect gravimetric methods are usually less precise than direct ones

In an indirect volumetric method a measured quantity of reagent is added in excess of the amount required to react with the desired constituent The excess reactant is then determined by titration and the amount of it reacting with the desired constituent is determined by difference

Methods of quantitative analysis depend greatly upon the nature of the substance being analyzed For this reason compilations of tested methods have been prepared each covering a certain type of material These compilations are available as reference books and laboratory manuals and cover such diverse fields as the analysis of steels nonferrous alloys foods minerals and ores gases technical products and agricultural products

Methods of quantitative analysis vary with the amount of sample taken and of constituent being determined A macroanalysis (decigram analysis) uses a sample of about 0 1 0 5 gram (g) The analytical balance and volumetric instruments are de-

)

s

o

very ... respec

A microanalysis (milligram analysis) uses a sample of about 1 10 mg and the balance is designed to read to a precision of 0 001 mg (1 microgram 1 μg or 1γ) The balance differs from the conventional one in being more delicately constructed of very light material It is necessary to have it on a firm vibration proof foundation, and great precautions are necessary to maintain draft free air at constant temperature and humidity See BALANCE (ANALYTICAL)

The instruments and apparatus used in microanalytical work are to some extent miniature replicas of macro instruments and apparatus but in some cases quite different techniques and apparatus are used Filtration for example is unlike that on a macro scale (see GRAVIMETRIC ANALYSIS) It is done by inserting into the vessel containing a suspension of the precipitate a filter stick consisting of a glass tube one end of which is closed by a porous sintered glass disk The filter stick and vessel have previously been weighed together and are treated as a unit Filtration is upward and is carried out by applying gentle suction to the tube After appropriate washing of the precipitate the filter stick and vessel are dried and again weighed together

An ultramicroanalysis (microgram analysis) uses a sample of approximately 0 001 mg A special quartz fiber torsion balance is used It has a capacity of about 20 mg and weighs to about 0 02 μg

Calibration of method Since the accuracy of a quantitative analysis depends in part on the nature of the material and on the nature and quantities of foreign constituents present calibration of methods of unknown applicability is often necessary

In the correction factor method the procedure is applied to a sample containing all the constituents in the given sample ...

Any r

rectio

known

In the synthetic sample method the procedure of analysis is applied in parallel to the unknown and to a sample containing constituents of the same nature and approximate amounts as those in the unknown and with the desired constituent added in known amount Any discrepancy in the value obtained on the synthetic sample is applied to correct the analysis of the unknown In a variation of this method the proposed procedure is applied to a sample of similar composition obtained from the National Bureau of Standards. Samples available include various ferrous and nonferrous alloys and many ores and minerals The values for the certified samples have been obtained by experienced analysts applying independent methods and using every precaution to ensure the highest degree of accuracy If concordant results are obtained on the standard sample by the proposed method it can be assumed to be reliable

Calibration of apparatus Analytical weights volumetric glassware and all other instruments that furnish numerical data should be calibrated

In an ordinary quantitative analysis if the same set of weights is used throughout it is necessary only to determine the relative masses of the weights That is the mass of the 10-g weight should be exactly twice that of the 5 g weight Similarly

Degrees of freedom	*t values*	*Degrees of freedom*	*t values*
1	12 71	15	2 13
2	4 30	16	2 12
3	3 18	17	2 11
4	2 78	18	2 10
5	2 57	20	2 09
6	2 45	22	2 07
7	2 36	24	2 06
8	2 31	26	2 06
9	2 26	28	2 05
10	2 23	30	2 04
11	2 20	40	2 02
12	2 18	60	2 00
13	2 16	120	1 98
14	2 14	∞	1 96

F test of significance. When sampling by variables, it is sometimes desirable to compare sample variances (squared standard deviations), S_1^2 and S_2^2, obtained from a pair of random samples of n_1 and n_2 observations from a population A sample variance is computed by either of the two formulae

$$S^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{\Sigma X^2 - (\Sigma X)^2/n}{n - 1}$$

To determine whether one sample variance is significantly larger than the other, an F ratio of the larger variance to the smaller is formed

$$F = \frac{S_1^2}{S_2^2}$$

If F is no larger than the F value found in Table 2 for $n_1 - 1 = \nu_1$ and $n_2 - 1 = \nu_2$ degrees of freedom, the two variances are not significantly different

If, on the other hand, F is larger than the tabled value, the variances are significantly different because such a difference can occur as a result of the inherent sampling variation only 1 time in 20 by chance alone See ANALYSIS OF VARIANCE, INDUSTRIAL ENGINEERING, INSPECTION AND TESTING [J MN]

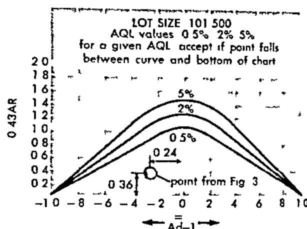


Fig 4 Curves for determining acceptance or rejection of inspected lots when using variables sampling

Bibliography A H Bowker and H P Goode *Sampling Inspection by Variables,* 1952, E L Grant, *Statistical Quality Control,* 2d ed, 1952 W G Ireson, *Sampling Tables for Inspection by Variables,* Stanford Univ, Appl Math Stat Lab Tech Rep 7, 1952, W G Ireson and G J Resnikoff, *Sampling Tables for Variables Inspection Based on the Range,* Stanford Univ, Appl Math Stat Lab Tech Rep 11 1952, J M Juran (ed) *Quality Control Handbook,* 1951, *Military Standard 105 A,* 1950

Quantitative chemical analysis

That branch of analytical chemistry which deals with the determination of the relative proportions of constituents in a compound or of components in a mixture The subject is divided into the broad fields of gravimetric methods, volumetric methods (liquid), gas volumetric methods optical methods electrical methods, and miscellaneous physico-

Table 2 Values of F for indicated degrees of freedom and probability of 0 05

ν_2 \ ν_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18 5	19 0	19 2	19 2	19 3	19 3	19 4	19 4	19 4	19 4	19 4	19 4	19 5	19 4	19 5	19 5	19 5	19 5	19 5
3	10 1	9 55	9 28	9 12	9 01	8 94	8 89	8 84	8 81	8 79	8 74	8 70	8 66	8 64	8 62	8 59	8 57	8 55	8 53
4	7 71	6 94	6 59	6 39	6 26	6 16	6 09	6 04	6 00	5 96	5 91	5 86	5 80	5 77	5 74	5 72	5 69	5 66	5 63
5	6 61	5 79	5 41	5 19	5 05	4 95	4 88	4 82	4 77	4 74	4 68	4 62	4 56	4 53	4 50	4 46	4 43	4 40	4 36
6	5 99	5 14	4 76	4 53	4 39	4 28	4 21	4 15	4 10	4 06	4 00	3 94	3 87	3 84	3 81	3 77	3 74	3 76	3 66
7	5 59	4 74	4 35	4 12	3 97	3 87	3 79	3 72	3 68	3 63	3 57	3 51	3 44	3 41	3 38	3 34	3 30	3 27	3 23
8	5 32	4 46	4 07	3 84	3 69	3 58	3 50	3 44	3 39	3 35	3 28	3 22	3 15	3 12	3 08	3 04	3 00	3 9	2 93
9	5 12	4 26	3 86	3 63	3 48	3 37	3 29	3 23	3 18	3 14	3 07	3 01	2 94	2 90	2 86	2 82	2 79	2 75	2 71
10	4 96	4 10	3 71	3 48	3 32	3 22	3 14	3 07	3 02	2 98	2 91	2 84	2 77	2 74	2 70	2 66	2 62	2 58	2 54
12	4 75	3 88	3 49	3 26	3 11	3 00	2 91	2 85	2 80	2 75	2 69	2 62	2 54	2 51	2 47	2 43	2 38	2 34	2 30
15	4 54	3 68	3 29	3 06	2 90	2 79	2 71	2 64	2 59	2 54	2 48	2 40	2 33	2 29	2 25	2 20	2 16	2 11	2 07
20	4 35	3 49	3 10	2 86	2 71	2 60	2 51	2 45	2 39	2 35	2.28	2 20	2 12	2 08	2 04	1 99	1 95	1 90	1 84
24	4 26	3 40	3 01	2 78	2 62	2 51	2 42	2 36	2 30	2 25	2 18	2 11	2 03	1 98	1 94	1 89	1 84	1 79	1 73
30	4 17	3 32	2 92	2 69	2 53	2 42	2 33	2 27	2.21	2 16	2 09	2 01	1 93	1 89	1 84	1 9	1 74	1 68	1 62
40	4 08	3 23	2 84	2 61	2 45	2 34	2 25	2 18	2 12	2 08	2 00	1 92	1 84	1 79	1 74	1 69	1 64	1 58	1 51
60	4 00	3 15	2 76	2 52	2 37	2 25	2 17	2 10	2 04	1 99	1 92	1 84	1 75	1 70	1 65	1 59	1 53	1 47	1 39
120	3 92	3 07	2 68	2 45	2 29	2 16	2 09	2 02	1 96	1 91	1 83	1 75	1 66	1 61	1 55	1 49	1 43	1 35	1 25
∞	3 81	3 00	2 60	2 37	2 21	2 10	2 01	1 94	1 88	1 93	1 75	1 67	1 57	1 52	1 46	1 39	1 32	1 22	1 00

chemistry is a form of applied mechanics Isaac Newton was aware of this and throughout the nineteenth century chemists tried to find mechanical explanations for chemical observations About 1900 however it became clear that the laws of mechanics which Newton had deduced from the behavior of macroscopic objects are not entirely valid on the atomic and subatomic scale The search for a correct set of basic mechanical principles for atoms culminated in 1926 with the discovery of quantum mechanics by W Heisenberg and E Schrodinger It has since become firmly established that the chemist must turn to this form of mechanics in order to find the mechanical basis for his observations *See* QUANTUM MECHANICS

This entails a fundamental change in point of view Certain concepts commonly used in dealing with macroscopic events must be modified or abandoned in discussing molecular and atomic phenomena In particular because of the uncertainty principle it is necessary to give up the notion of orbits or definite paths along which particles, especially electrons are predestined to move (*see* UNCERTAINTY PRINCIPLE) The state of a system is instead described by means of statistical concepts The most important of these is the probability distribution function which describes the relative probabilities of the different possible spatial configurations of the electrons and nuclei out of which atomic systems are constructed The probability distribution in turn is directly related to the wave function of the state of the system The laws of quantum mechanics as they are commonly used in quantum chemistry consist of rules for finding the possible wave functions of chemical systems and for calculating the properties of the systems from these wave functions

Because of mathematical difficulties it is rarely possible to apply the rules of quantum mechanics precisely to chemical systems In order to make any progress it has become necessary to resort to approximate methods Most of the successful concepts of quantum chemistry are formulated in the framework of these approximate methods As a result there are now qualitative and semiquantitative quantum mechanical explanations for a great many chemical phenomena Most of these phenomena could never have been understood by use of Newtonian mechanics alone The vocabulary of the quantum chemist is consequently filled with nonclassical terms which have no counterpart in macroscopic systems

Schrödinger equation The most important basic postulate of quantum mechanics may be stated in the following simplified form Consider a particle of mass m wh [illegible] dimension [illegible] given by th[illegible] ing the part [illegible] e narrow region between x and $x + dx$ is $P(x)\,dx$ The function $P(x)$ is the probability distrib[illegible] (se[illegible] wav[illegible]

$$P(x) = [\psi(x)]^2 \qquad (1)$$

The wave function is determined by solving Schrödinger's differential equation

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2 m}{h^2}[E - V(x)]\psi = 0 \qquad (2)$$

where E is the total energy of the system $V(x)$ is the potential energy when the particle is at x and h is Planck's constant ($6\,625 \times 10^{-27}$ erg sec) If a so-called energy operator (or Hamiltonian operator) is defined as

$$H = -(h^2/8\pi^2 m)\; d^2/dx^2 + V(x) \qquad (3)$$

then the Schrödinger equation can be written more concisely as

$$H\psi = E\psi \qquad (4)$$

See HAMILTON'S EQUATIONS OF MOTION

For more complex systems containing n particles moving in more than 1 dimension the Hamiltonian operator takes the form

$$H = -(h^2/8\pi^2)\sum_{\substack{\text{all}\\ \text{particles}}}\frac{1}{m_i}\nabla_i^2 + V(x_1,y_1\,z_1,\ \ ,z_n) \qquad (5)$$

where m_i is the mass of the ith particle, ∇_i^2 is the Laplacian operator for the ith particle,

$$\nabla_i^2 = \partial^2/\partial x_i^2 + \partial^2/\partial y_i^2 + \partial^2/\partial z_i^2 \qquad (6)$$

x_i, y_i and z are the cartesian coordinates of the ith particle and $V(x_1\,y_1\,z_1\ \ ,z_n)$ is the potential energy when the particles have the specified cartesian coordinates The wave function, $\psi(x_1\,y_1\,z_1\ \ldots,z_n)$, is now a function of all of the coordinates, it is found by solving Eq (4), where H is defined by Eq (5) For a more detailed discussion of Schrodinger's equation *see* QUANTUM THEORY, NONRELATIVISTIC.

The phenomenon of energy quantization arises because physically meaningful solutions of Schrödinger's Eq (4) can be found for many systems only if the energy E has certain distinct values One of the primary aims of quantum chemistry is to find these permitted values of the energy for molecular systems because these energies are generally of much greater interest to the chemist than are either the wave functions or the probability distributions It is found however that the Schrödinger partial differential equations which arise from even the simplest chemical systems are much too complicated to be solved precisely by available mathematical methods Exact values of the permitted energies therefore cannot be found

Variation method Consequently it is necessary to proceed by using approximate methods for finding the wave functions and their corresponding energies The most useful of these methods the variation method has produced most of the important quantum chemical concepts This method depends upon the following theorem (the variation principle)

the others should be in corresponding proportion to the 10 g weight The mass of one of the weights is assumed to be correct as marked on the weight and the correction factor to be applied to each of the other weights to maintain the exact theoretical proportion is determined by experiment It is more desirable however, to use as a standard of refer ence a weight of known mass such as one certified by the Bureau of Standards

In calibrating volumetric glassware the weight of water (or mercury) at a given temperature con tained in, or delivered by, the apparatus is found From the known density of the liquid at the given temperature, the true volume can be calculated and compared to the volume indicated by the mark ing on the volumetric apparatus The fundamental standard is the liter which is the volume occupied by 1 kilogram (kg) of water at the temperature of its maximum density (approximately 4°C) The normal temperature for calibrating volumetric glassware in the United States is 20°C, which means that to contain a true liter, a flask must be so marked that at 20°C its capacity will be equal to the volume of water which at 4°C weighs 1 kg in vacuo

Several types of calibration are encountered in other branches of analytical chemistry A com mon calibration procedure in colorimetry is to bracket the unknown between two standards at known concentration just above and just below that of the unknown Another is useful in emis sion spectroscopy where the density of spectral lines is a quantitative measure of concentration In order to eliminate the effects of undesirable variations in the excitation an internal standard method is used whereby the density of a spec tral line of the element of unknown concentration is compared to that of a line of an element of known concentration The pairs of lines called homologous pairs respond in the same way to changes in excitation conditions A standard addi tion method is useful in cases where the nature of the sample may preclude the direct use of synthetic samples Here the slope of a calibration curve (a curve showing corrections to be applied under spec ified conditions of analysis) is determined from the relative positions on the curve of the value ob tained from the diluted sample and that obtained on an equal volume of sample to which has been added a standard solution containing a known ad ditional amount of the constituent being deter mined

Graphs are frequently used in analytical chemis try for many purposes A nomograph is a device by which the numerical result of a given calculation can be read directly from a previously drawn scale or series of scales It has the advantage over a slide rule in being equally applicable to calcula tions containing additive and subtractive terms *See* NOMOGRAPH Since a separate nomograph is needed for each formula to be solved nomographs are of practical use only when the same type of analytical calculation is made repeatedly

Other graphs encountered in quantitative analy sis are calibration corrections, tabulation of values derived from natural laws and titration curves in which pH values, conductances, or electrode po tential values are plotted against buret readings in order to establish equivalence points In such graphs some variables change in logarithmic fash ion giving sigmoidal curves, others change in arithmetic proportion and yield straight line graphs *See* ANALYTICAL CHEMISTRY, SPECTRO CHEMICAL ANALYSIS, SPECTROPHOTOMETRIC ANALY SIS, VOLUMETRIC ANALYSIS

[S C S]

Bibliography C W Foulk, H V Moser and W M MacNevin, *Quantitative Chemical Analysis* 1952, I M Kolthoff and E B Sandell *Textbook of Quantitative Inorganic Analysis*, 3d ed 1952 W W Scott and N H Furman (eds) *Standard Methods of Chemical Analysis* 2 vols 5th ed, 1939

Quantization

A term referring to the fact that certain observable quantities have only a discrete set of allowed val ues, for example the allowed values of the orbital angular momentum of a particle are

$$\sqrt{l(l+1)}\,h/2\pi$$

where $l = 0$ or a positive integer and h is Planck's constant Quantization is the process of obtaining these values The existence of such quantized ob servables is well established, the quantization rules determining the allowed values or eigenvalues of each observable are predicted by quantum me chanics *See* EIGENVALUE, QUANTUM MECHANICS, QUANTUM THEORY NONRELATIVISTIC [E G]

Quantum (physics)

An indivisible quantity of electromagnetic energy For a light wave of frequency f, the quantum of en ergy is hf, where h is Planck's constant The term light quanta often is used interchangeably with photons, the massless particles which according to present theory transport the energy and momen tum of a light wave (*see* PHOTON) The import of the existence of quanta is that the energy in a light wave changes discontinuously in multiples of hf Similar discontinuities characterize quantum me chanics the modern theory of matter *See* QUAN TUM MECHANICS, QUANTUM THEORY NONRELATIV ISTIC [E G]

Quantum chemistry

A branch of physical chemistry concerned with the explanation of chemical phenomena by means of the laws of quantum mechanics Starting with the time of John Dalton physical and chemical re search in the nineteenth century demonstrated the essential truth of the atomic theory—that chemical change and the properties of matter result from the interactions and motions of atoms It follows from the atomic theory that chemistry is a consequence of the science of the motions of particles, that is

ratio of the parameters A and B and not on their individual values This is a consequence of the so-called linear character of the Hamiltonian operator The four integrals H_{aa} H_{ab} I, and S can be evaluated readily but it is sufficient here to state that both H_{aa} and H_{ab} are negative and that both I and S are positive To find the value of the ratio p which minimizes W, the minimax *condition* $dW/dp = 0$ gives the result

$$p = \pm 1 \qquad (13)$$

Because H_{ab} is negative and S is positive inspection of Eq (12) shows that the root $p = +1$ leads *to* a minimum value of W and is the optimum value of this parameter

This leads to the approximate wave function for the ground state of H_2^+

$$\phi = A(1s_a + 1s_b) \qquad (14)$$

and the approximate energy

$$W = \frac{H_{aa} + H_{ab}}{I + S} \qquad (15)$$

Evaluation of W at various values of the interproton distance R shows that W passes through a minimum value at $R = 1.32$ A the value of W at this separation corresponds to a dissociation energy of 1.76 ev Although it is not in very close quantitative agreement with the experimentally observed values mentioned previously this calculation is significant because the existence of a stable H_2 molecule cannot be explained even qualitatively either by classical mechanics or by the well known quantum modifications of classical mechanics introduced by Niels Bohr

Closer inspection of the integrals H_{aa} and H_{ab} in Eq (15) reveals that in this approximate treatment all of the stability of H_2^+ arises from the integral H_{ab} This means that if one had chosen as the trial wave function

$$\phi = \exp(-r_a/a_0) \qquad (16)$$

which leads to the estimated energy

$$W = H_{aa}/I \qquad (17)$$

the calculation would not have led to the prediction that the H_2 molecule would be stable Therefore an important element in strengthening the bond in H must be the fact that the electron is able to move from one proton to the other This factor is called exchange and the integral H_{ab} which results from it is called the exchange integral The integral H_{aa} is called the coulombic integral It is clear that the concept of exchange can have meaning only in connection with the particular type of approximate wave function used in Eq (14) yet the concept provides a valuable although relatively qualitative means of understanding the origin of chemical bonding

It should be mentioned that a more physical explanation of the success of trial wave function Eq (14), is to be found in the fact that it gives a considerably higher probability for finding the electron between the two protons than does trial function Eq (16) The exchange integral H_{ab} is merely the mathematical manifestation of this increase in charge density near the midpoint of the bond

Hydrogen molecule A similar approach may be used for the hydrogen molecule Here there are two electrons which may be denoted by the numbers 1 and 2 as well as two protons denoted by a and b The trial wave function may be written

$$\phi = 1s_{a1}1s_{b2} \qquad (18)$$

which signifies that electron 1 is to be found in a 1s hydrogenic orbital in the vicinity of proton a, and electron 2 is in a similar orbital centered on proton b Substitution of this trial function into Eq (7), using the Hamiltonian operator appropriate to the hydrogen molecule does not however give any indication of the observed strong chemical bond in the hydrogen molecule On the other hand W Heitler and F London showed that if the two electrons are allowed to exchange positions between the two protons by writing the trial wave function

$$\phi = 1s_{a1}1s_{b2} + 1s_{a2}1s_{b1} \qquad (19)$$

then a bond between the two protons is predicted which has a strength comparable to the observed bond strength in H_2

Resonance The concept of resonance arises from these treatments of H_2^+ and H_2 One can say that in both of these molecules extra stability is obtained because the system can exist in more than one structure The stabilization of benzene and other aromatic molecules can be explained in the same way because the valence bonds in these molecules can be drawn in several different ways The importance of this resonance concept in chemistry can hardly be [illegible]

pr[illegible]
of [illegible] A and B (for the Heitler London theory of valence see VALENCE) If an electron on an isolated A atom has the wave function ψ_A and one on an isolated B atom has the wave function ψ_B then an approximation to the wave function of the two electrons involved in the A—B bond would be

$$\phi = \psi_A(1)\psi_B(2) + \psi_A(2)\psi_B(1) \qquad (20)$$

If the electron affinities and ionization energies of A and B are markedly different the trial wave function may be improved by adding to this trial function terms corresponding to the ionic structures A^+B^- and A^-B^+

$$\phi = \psi_A(1)\psi_B(2) + \psi_A(2)\psi_B(1) + C\psi_B(1)\psi_B(2) + D\psi_A(1)\psi_A(2) \qquad (21)$$

where C and D are variable parameters which are selected so as to minimize the energy The importance of the ionic terms in bringing additional sta

Let H be the Hamiltonian operator for a system, and let ϕ represent some approximation to the wave function of the state of lowest energy of the system (ground state) Then the quantity

$$W = \frac{\int \phi H \phi \, dV}{\int \phi^2 \, dV} \qquad (7)$$

is an approximation to the correct energy, E_0, of the ground state *Furthermore, the estimated energy, W, is always algebraically larger than the correct energy, E_0* (that is, $W - E_0 > 0$) Only if one happens to have chosen the correct wave function ψ_0 for the ground state, so that $\phi = \psi_0$, will it happen that $W = E_0$, in this case $H\phi = E_0\phi$, so that

$$\int \phi H \phi \, dV = E_0 \int \phi^2 \, dV$$

The quantity dV in Eq (7) represents the volume element $dV = dx_1, dy_1, dz_1, \ldots, dz_n$, and the integration is carried out over all of the values of the coordinates accessible to the system

The variation principle is utilized by selecting some intuitively reasonable approximate trial wave function, ϕ, whose shape can be changed in a systematic fashion by altering the values of constant parameters that appear in the function The set of parameters that leads to the smallest possible value of W *is then found According to the variation principle*, this minimum value of W is the closest to the correct value E_0, for the energy of the ground *state that it is possible to get with a trial function* of the initially assumed form Furthermore, when the parameters which minimize W are inserted into ϕ, one obtains a closer approximation to the correct wave function than can be found with any other set of parameters

The success of this method depends of course on the ability of the quantum chemist to guess at trial functions which are good approximations and at the same time contain parameters in such a form that W can be minimized without undue labor Intuition and previous experience play a major role here Progress in quantum chemistry since World *War II has been much concerned with devising new* and more tractable forms for the trial wave function, and with developing means of dealing with more complex types of functions High speed computing machines have made great contributions in this effort At the present time fairly good calculations of some of the lower energy levels of diatomic molecules such as oxygen and nitrogen appear to be possible

A few simple examples will show at the same time how the *variation method is utilized* and how important insights into the mechanical basis of chemical behavior may be obtained A system con*sisting of two protons* fixed in space at a distance R from each other, with a single electron which is able to move about in their vicinity, represents the hydrogen molecular ion, H_2^+, which is known to exist in hydrogen electric discharges, and which is the simplest of all molecules and offers the most elementary example of a chemical bond Exp mental observations of the spectra of hydrogen charge tubes show that in its state of lowest ene the two protons in H_2^+ are separated by 1.0 and that the dissociation into a proton and a hy gen atom ($H_2^+ \rightarrow H^+ + H$) requires an energ 2 791 ev

Hydrogen molecular ion. The exact wave f tion for the ground state of the hydrogen ato $\psi = \exp(-r/a_0)$, where r is the distance betw the proton and the electron and a_0 is a cons *length, known as the Bohr radius equal* $0\ 52917 \times 10^{-8}$ cm One might expect that in the electron will move about both protons in so *what the same way that it moves in a hydr* atom, so it is reasonable to write as the trial w function,

$$\phi = A \exp(-r_a/a_0) + B \exp(-r_b/a_0) \quad ($$

where A and B are constant parameters whose ues will be selected so as to minimize W, r_a is distance from the electron to one of the prot and r_b is the distance to the other proton Beca the wave function for hydrogen in its ground s is usually indicated by the symbol 1s, it is conv ent to write $1s_a$ for $\exp(-r_a/a_0)$ and $1s_b$ for $(-r_b/a_0)$, so that

$$\phi = A1s_a + B1s_b$$

For H_2^+, the Hamiltonian operator is

$$H = -(h^2/8\pi^2 m)\nabla^2 - (e^2/r_a) - (e^2/r_b) + (e^2/R)$$

where e is the electronic charge, $-(e^2/r_a)$ (e^2/r_b) is the potential energy due to the elec static interactions of the electron with the two p tons (Coulomb's law), *and e^2/R represents* electrostatic repulsion of the protons Substitut Eqs (8) and (9) into Eq (7) gives for the proximate energy

$$W = \frac{A^2 H_{aa} + B^2 H_{bb} + AB(H_{ab} + H_{ba})}{(A^2 + B^2)I + 2ABS} \quad ($$

where
$$H_{aa} = \int 1s_a H 1s_a \, dV \quad (1$$
$$H_{bb} = \int 1s_b H 1s_b \, dV \quad (1$$
$$H_{ab} = \int 1s_a H 1s_b \, dV \quad (1$$
$$H_{ba} = \int 1s_b H 1s_a \, dV \quad (1$$
$$I = \int 1s_a^2 \, dV \quad (1$$
and
$$S = \int 1s_a 1s_b \, dV \quad (1$$

Because the two protons are identical it must true that

$$H_{aa} = H_{bb} \quad \text{and} \quad H_{ab} = H_{ba} \quad ($$

so that

$$W = \frac{(A^2 + B^2)H_{aa} + 2ABH_{ab}}{(A^2 + B^2)I + 2ABS} = \frac{(1 + \rho^2)H_{aa} + 2\rho H_{ab}}{(1 + \rho^2)I + 2\rho S} \quad (1$$

where $\rho = B/A$ Evidently W depends only on th

to remember that even at absolute zero temperature the electromagnetic field is present because of the zero point fluctuations of the oscillators These zero point oscillations are called vacuum fluctuations Further when the field is treated quantum mechanically it is possible to understand the classically incomprehensible phenomenon of the external photoelectric effect Thus when light shines on a metal for example electrons are freed from the surface by absorbing energy from the electromagnetic field The spectrum of energy of the ejected particles is independent of the intensity of the light More intense beams produce more electrons but the relative number of electrons with various energies remains fixed This result is easily understood quantum mechanically by regarding the interaction between the electrons and the electromagnetic field as produced by the absorption of discrete quanta of energy from the quantized oscillators of the electromagnetic field These discrete

the theory involves consists mainly of electrons Electrons of low energy may be treated consistently as particles according to the principles of quantum mechanics at low energies their wave functions satisfy the Schrodinger equation At higher energies it is necessary to consider them as the quanta of a field which obeys the relativistic equation of P A M Dirac (For a statement and discussion of this equation *see* QUANTUM THEORY RELATIVISTIC) For simplicity consider the consequence of these facts for the spectrum of energy states of the one-electron hydrogen atom In the nonrelativistic description the nth excited state is actually a multiplet of states (it is degenerate) To a large extent the multiplicity can be said to be accidental By

thi

for

ang

per

ene for each value of the orbital angular momentum $l = 0$ $n - 1$ where n is the total quantum number (*see* QUANTUM NUMBERS) Each of these levels is in turn a multiplet of $2(2l + 1)$ states because of the so-called spatial degeneracy of the orbital and spin angular momenta However in the more accurate description provided by the Dirac equation the degeneracy is reduced that is, each level is accidentally only a multiplet of two states The splitting of the multiplets may be regarded as produced by the relativistic corrections The spectrum of states predicted by the Dirac theory is much closer to the observed spectrum than that predicted by the Schrodinger theory

The Dirac equation as well as leading to the concept of electrons as quanta of the field also predicts for the field the existence of positively charged quanta of the same mass as electrons These are called positrons The observation of these latter quanta was consequently a triumph of the field theory of the electrons The Dirac field also can be in thermal equilibrium with the walls of a container, but the manifestations are somewhat different from the case of the electromagnetic field for three reasons First the quanta of the Dirac field are charged, so that the field fluctuations are constrained by the law of charge conservation Second, the quanta of the Dirac field have a rest mass which is quite large, so quanta of the field can be created and destroyed only at extreme temperatures At low temperatures the system behaves just like a gas of a fixed number of particles At the extremely high temperatures found in stars, the field aspects are the important ones and it is no longer useful to regard the system as an ensemble of particles Finally and most important the electrons obey the Pauli exclusion principle, so that no more than one particle with a given set of quantum numbers may be present For electrons confined in a metal these restrictions lead to the well known Fermi distribution of electron energies

Interacting fields The coupled Dirac (electron) and Maxwell (electromagnetic) fields represent the complete system which is governed by quantum electrodynamics The coupling as well as permitting the emission and absorption of light by the electrons and the production of pairs of electrons and positrons by the light (photons), also modifies the description of the electron because speaking loosely, the electron as it moves interacts with the vacuum fluctuations of the electromagnetic field In like manner the

tons)

vacuu

action vacuum fluctuations are sometimes called for the

equations In both cases the effect of the interactions with the virtual quanta produces only a small change in the already mentioned properties of the uncoupled fields This is because the coupling which is measured by the so called fine structure constant $\alpha = 2\pi e^2/hc$ (e is the electron charge h is Planck s constant c is the velocity of light) is relatively weak The properties of the electromagnetic field alone (in the absence of real charged matter) are only slightly modified by the interaction with the virtual charged matter, or vacuum pair fluctuations The equations for the electromagnetic field now become in effect slightly nonlinear

it which annihilate to yield pho

bility to a bond depends on the difference in the electronegativities of the atoms A and B

The functions ψ_A and ψ_B are known as atomic orbitals and a considerable improvement in the wave function can be made by allowing the shapes of these orbitals as they exist in molecules to differ from their shapes in the free atoms Generally it is desirable to use atomic orbitals which tend to concentrate electrons in the region of the chemical bond that is to be formed This change in shape of the atomic orbitals is conveniently brought about by adding to a given atomic orbital the orbitals of other states of the atom a procedure known as hybridization This procedure leads to a simple and plausible explanation for many basic observations in the field of stereochemistry

An alternative approach to the chemical bond is provided by the molecular orbital method of writing approximate wave functions for molecules This may be illustrated with the hydrogen molecule Let ψ represent the exact wave function for the H_2^+ molecule (or a good approximation to it) Then the molecular orbital approximation to the wave function of H is

$$\phi - \psi(1)\psi(2) \quad (22)$$

where $\psi(1)$ is the wave function of H_2^+ containing the coordinates of electron number 1 and $\psi(2)$ is the same function containing the coordinates of electron number 2 If one makes the approximation

$$\psi = 1s_a + 1s_b \quad (23)$$

this gives from Eq (22)

$$\begin{aligned}\phi &- (1s_{a1} + 1s_{b1})(1s_{a2} + 1s_{b2}) \quad (24)\\ &= 1s_{a1}1s_{b2} + 1s_{a2}1s_{b1} + 1s_{a1}1s_{a2} + 1s_{b1}1s_{b2}\end{aligned}$$

The first two terms here are the same as trial function Eq (19) and the last two terms represent ionic structures Because of the overemphasis on the ionic terms this trial function does not give as good results quantitatively as Eq (19) If however additional terms are added to the trial function in which the electrons are placed in excited orbitals of H_2 the numerical results come closer to the observed energies than do the results with Eq (19) This procedure is known as configuration interaction In complex molecules the molecular orbital method (including configuration interaction) is more easily carried out and gives better results than calculations based on Eq (20)

It is evident that the variation method provides the chemist with many conceptual aids to the understanding of the origin of chemical bonds and their properties The shrewd combination of these conceptual aids with empirical observations has made it possible to deduce a useful quantitative theory of chemical bonds and their variations in different molecules Nevertheless *the trial functions* described above are far from accurate descriptions of the true wave functions of molecules When better and more complex trial functions are used the concepts employed in interpreting simple trial functions lose much of their meaning

The discussion given above refers to the treatment of the ground states of molecules The same approach can however be extended to the excited states of molecules (although the results tend to be even less accurate and the calculations more difficult) In this way a quantum chemical basis has been developed for the phenomena of light absorption color and other optical properties of matter *See* CHEMICAL BINDING CHEMICAL STRUCTURES MOLECULAR STRUCTURE AND SPECTRA RESONANCE (MOLECULAR STRUCTURE) [W.J.K.]

Bibliography W Kauzmann *Quantum Chemistry An Introduction* 1957

Quantum electrodynamics

That part of quantum field theory which covers the electromagnetic field and its interaction with electrically charged matter or matter fields (*see* QUANTUM FIELD THEORY) Broadly speaking the theory includes all of atomic physics since the structure and interaction of atoms are governed by electrical forces However because the main features of the atomic system are accounted for by the electrostatic Coulomb force between charged particles and because the treatment of this force does not require any detailed use of electromagnetic theory in its usual context quantum electrodynamics refers to the refinements of the description of the interactions which are necessary to account for the fine structure of atomic systems To be specific the fine structure of the atomic system that is produced by the quantum fluctuations of the electromagnetic and matter fields and the interaction of light (electromagnetic fields) with charged matter is the main province of quantum electrodynamics The system which quantum electrodynamics describes is thus the electromagnetic field and charged matter

Quantum electrodynamics was formulated by P A M Dirac and by W Heisenberg and W Pauli However in its initial version there were numerous ambiguities Later J Schwinger and R P Feynman independently showed how to formulate the theory in such a way as to avoid these ambiguities

It is convenient to divide the subject matter of quantum electrodynamics into three parts the free electromagnetic field the free charged matter and interaction between the two

Free electromagnetic field In classical physics the electromagnetic field is described by the Maxwell equations However the classical theory does not properly account for the observed spectrum of black body radiation (that is the distribution of energy with respect to frequency in the electromagnetic field enclosed by walls in thermal equilibrium with the field at a given temperature) However if the field is quantized *that is represented* by a superposition of quantum mechanical oscillators (which also obey Maxwell's equations) the energy spectrum is correctly given It is important

pace See ELECTROMAGNETIC RADIATION, FIELD THEORY CLASSICAL GRAVITATION

In addition to these fields which are familiar on ... cosmos ... lassical ... inds its

pplication to the understanding of the dynamical properties of the so-called elementary particles (electrons protons etc) which are the fundamental constituents of matter (see ELEMENTARY PARTICLE) The reason that quantum fields are suitable for the description of such particles is found in two facts First elementary particles of a given kind may be produced in profusion *because* of the interconvertibility of *energy and* matter Thus in the interaction of atomic matter particles *may be created* and *destroyed Second because of* the quantum mechanical wave interference properties of matter the classical notion of a field enters In fact in the notion *of a* quantum field the wave particle or better field particle duality characteristic of matter finds its most perfect expression

The theories of such fields must be formulated so that the "kinematical *properties are consistent* with the principles *of relativity* the dynamics is governed by *quantum* mechanics The subsequent discussion is based on the assumption that the *reader* is somewhat familiar with the subject matter of relativity and quantum mechanics See QUANTUM MECHANICS *QUANTUM THEORY* NONRELATIVISTIC *QUANTUM THEORY* RELATIVISTIC RELATIVITY RELATIVISTIC MECHANICS

Scalar field The particle properties just described emerge from a continuous field when it is treated quantum mechanically This can be shown in an elementary way by using the so-called scalar field as an example The scalar field is an invariant *function of the* space-time coordinates $\mathbf{x}$ and t When it describes free particles they must conform with the relativistic connection between the energy and momentum of a free particle of mass m that is $E^2/c^2 = \mathbf{p}^2 + (mc)^2$ where E is the energy $\mathbf{p}$ the momentum and c the velocity of light According to the general quantum mechanical rule

$$E \to i\hbar\frac{\partial}{\partial t} \qquad \mathbf{p} \to \frac{\hbar}{i}\nabla$$

are the expressions for E and $\mathbf{p}$ acting on functions of space and time, so the "field equation for ϕ is

$$\left[i\hbar\frac{\partial}{\partial(ct)}\right]^2\phi = \left[\left(\frac{\hbar}{i}\nabla\right)^2 + (mc)^2\right]\phi$$

If a Fourier transformation of the spatial dependence is made the equation

$$\left[i\hbar\frac{\partial}{\partial(ct)}\right]^2\phi = \left[\mathbf{p}^2 + (mc)^2\right]\phi$$

is obtained For each value of $\mathbf{p}$ this equation describes a harmonic oscillator of frequency $\{[\mathbf{p}^2 + (mc)^2]^{1/2}c\}/\hbar$ and because the oscillator has the energy spectrum $(n + \frac{1}{2})[\mathbf{p}^2 + (mc)^2]^{1/2}c$ where $n = 0\ 1\ 2$ the energy levels of the dynamical *system* described by the scalar field ϕ are $E = \Sigma[n(\mathbf{p}) + \frac{1}{2}][\mathbf{p}^2 + (mc)^2]^{1/2}c$ where for each value of $\mathbf{p}$ the possible values of $n(\mathbf{p})$ are integers (see HARMONIC OSCILLATOR) Thus the changes in the energy values occur in multiples of $[(mc)^2 + \mathbf{p}^2]^{1/2}c$ and the spectrum can be interpreted by saying that there are $n(\mathbf{p})$ *particles or quanta of energy* $[(mc)^2 + \mathbf{p}^2]^{1/2}c$ The zero point energy of each oscillator, $\frac{1}{2}[\mathbf{p}^2 + (mc)^2]^{1/2}c$ may for convenience be omitted since the state of lowest energy when $n(\mathbf{p}) = 0$, which is called the vacuum can most conveniently be assigned zero energy However it should be emphasized that this energy has its origin in the fluctuations in the field in the state of lowest energy, the existence of these fluctuations is of great importance The field $\phi(\mathbf{p},t)$ is the coordinate of the oscillator and just as in the case of a single harmonic oscillator

$$\phi(\mathbf{p}\ t) \pm i\hbar\frac{1}{[\mathbf{p}^2 + (mc)^2]^{1/2}}\frac{\partial}{\partial(ct)}\phi(\mathbf{p}\ t) = \phi^{(\pm)}(\mathbf{p}\ t)$$

are the operators which when applied to an energy eigenstate respectively lower and raise the energy level by $[(mc)^2 + \mathbf{p}^2]^{1/2}c$ Hence $\phi^{(\pm)}$ are called particle-absorption and emission operators

Quantum electrodynamics The quantum field which has been studied in the greatest detail and the theory of which has had the greatest success in accounting for the observed properties of atomic systems is the same field which is most well known classically—the electromagnetic field For this reason it is considered separately (see QUANTUM ELECTRODYNAMICS)

Gravitational field The quantum mechanical treatment of the gravitational field (the other field familiar at a macroscopic level) has so far not been developed to the same extent and with the same success as that of the electromagnetic field There are two reasons for this First from a pragmatic point of view the gravitational field has so far had no relevance in the understanding of atomic phenomena because of the extremely weak coupling between this field and matter For example the ratio of the electrostatic force between the electron and proton in the hydrogen atom to the gravitational force is $\sim 10^{39}$ Further on the theoretical side, the treatment of the gravitational field quantum mechanically is extremely difficult because of the nonlinearity of the gravitational field equations and more significantly because the physical content of the gravitational theory is still not well understood The quantum of the gravitational field a theoretically deduced particle is called the graviton (see G...

tr ... gravita ... ra ... uut a quite ... cullent quantum field theories These are mentioned subsequently

tons again but with their momenta changed from the initial values A phenomenon which will be perhaps more easily observed is the so called Delbruck scattering In this case the photon is scattered by the virtual pairs which are polarized by the static Coulomb field of a nucleus

Lamb shift and g factor For the electron the effects produced by the coupling to the virtual photons are also striking and lead to a number of effects which have been accurately measured Agreement of the measurements with the detailed predictions of the theory is at present perfect First the electron as it moves interacts with the vacuum fluctuations of the electromagnetic field and this alters the inertia of the particle However the observed electron is always in the presence of the vacuum fluctuations and therefore the effect of these fluctuations on the mass cannot be directly observed that is the observed mass of the electron is the mass in the presence of the vacuum fluctuations of the electromagnetic field However when the electron is bound to an atom the effect on the inertia is slightly different and depends upon which excited state of the atom the electron occupies This leads to a splitting of the doubly degenerate states in hydrogen This splitting called the Lamb shift is largest in the case of the $2S_{1/2} - 2P_{1/2}$ states has been accurately observed and is in perfect agreement with the prediction of the theory at this time (For further information on the Lamb shift *see* ATOMIC STRUCTURE AND SPECTRA) Another effect is caused by the presence of the virtual pairs in the vacuum These are polarized by the electron s charge so that the charge of the electron is actually measured in the presence of the shield of the positively charged ends of the pairs Again this effect cannot be directly observed since all charges are always surrounded by such a shield However just as in the case of the mass the shield is altered when the electron is bound in an atom and the resulting effect on the energy levels has been included in the Lamb shift calculation which agrees well with the measured shift Further these vacuum interactions are altered when the electron moves in a magnetic field so that the electron s g factor (spin magnetic moment) comes out to be slightly different from that predicted by the Dirac theory The value predicted by quantum electrodynamics agrees perfectly with the observed value For a discussion of the numerical values involved in the g factor *see* ELECTRON SPIN

Positronium The best system for testing the predictions of quantum electrodynamics from the point of view of the theorist would be the atom called positronium formed by an electron positron pair However because of the process of pair annihilation such a system is unstable that is the electron and positron almost immediately combine to form photons Because of this very short lifetime it has thus far been impossible to produce enough positronium so that one can test the predictions of the theory to an accuracy which is better than that in the experiments already mentioned *See* POSITRONIUM

Effects of vacuum interactions In summary the predictions of quantum electrodynamics agree perfectly at present with all observations of the properties of electrons and light However there is one outstanding theoretical blemish which is at this time unresolved This has to do with two of the effects mentioned which cannot be directly observed, namely the effects of the vacuum interactions on the mass and on the charge of the free particle If one calculates these effects using the methods of perturbation theory they do not come out to be finite However all other effects calculated relative to the observed mass and observed charge are finite This defect is caused by the fact that in order to get a relativistically invariant (and in several other respects consistent) theory it is necessary to regard the electron as an object having zero radius This leads to the coupling of the charge to light of arbitrarily short wavelengths The interaction with vacuum fluctuations of these very high energy virtual quanta is most important for the mass and charge of the electron The present methods of handling the mathematics of the theory are unsatisfactory and lead to an infinite contribution to the mass of the particle coming from the high-energy virtual quanta Likewise in this method of calculation the shielding of the electron s charge by high energy virtual pairs comes out to be perfect, thus the observed charge of the electron should be zero Whether these defects are a consequence of the inadequacy of the approximations or are inherent inconsistencies in the method of formulating the theory is at present unknown [K A J]

Bibliography A I Akhiezer and V B Berestetsky *Quantum Electrodynamics* AEC TR 2876 1957 N N Bogoliubov and D V Shirkov *Introduction to the Theory of Quantized Fields* 1959 J M Jauch and F Rohrlich *The Theory of Photons and Electrons* 1955 J S Schwinger (ed) *Selected Papers on Quantum Electrodynamics* 1958 W E Thirring *Principles of Quantum Electrodynamics* 1958

Quantum field theory

The branch of physics which deals with the quantum mechanics of fields Quantum field theory represents a large fraction of the total effort expended by theoretical physicists

Fields are familiar in classical physics as systems which require an infinite set of variables for their description This infinite set is usually given as the values over all of space of a continuous function or set of functions which characterize some physical quantity Thus for example the classical Newtonian gravitational field is described by giving at each point in space the value of the gravitational potential energy Another classical field is the electromagnetic or Maxwell field which is determined by the electrical and magnetic forces a charged particle would experience everywhere in

difficulties are associated with the point structure of the particles which in the language of quantum field theory means that the couplings between the fields are local that is the field equations are differential equations and not integral equations When the attempt is made to solve such equations by expansions in the coupling constants one finds that the self energy and other such quantities are not finite Classically the self energy of a particle coupled to a field is the energy contained in the particle s own field In quantum theory a similar effect results from the fact that each particle field is coupled to the vacuum fluctuations of other fields The perturbation method calculation of such quantities depends strongly on the coupling of one field to the high energy fluctuations in the others. However it is possible by using the techniques of renormalized perturbation theory to calculate most quantities It is not known whether the infinities encountered in the perturbation solution of the field equations are a fundamental inconsistency built into all quantum field theories or whether they are simply a consequence of the perturbation method of solution *See* PERTURBATION (QUANTUM MECHANICS)

Symmetry laws Because in the case of strongly coupled fields no adequate method is known for solving the field equations other more modest approaches have been used These are based on making maximum use of the fundamental symmetries built into all physical theories One attempts to write for a given physical quantity the most general form consistent with all the symmetries built into the theory Instead of trying to calculate the quantity using the specific form of the theory an attempt is made to relate each such quantity to others In some cases this modest approach provides information about the nature of the fundamental interactions Eventually however the basic problem of the fundamental dynamics of elementary particles must be faced *See* SYMMETRY LAWS (PHYSICS) [K A J]

Bibliography E Fermi *Elementary Particles* 1951 S S Schweber H A Bethe and F De Hoffmann *Mesons and Fields* vol 1 1955 H Umezawa *Quantum Field Theory* 1956 G Wentzel *Quantum Theory of Fields* 1949 *see also* QUANTUM ELECTRODYNAMICS

Quantum mechanics

The modern theory of matter of electromagnetic radiation and of the interaction between matter and radiation also the mechanics of phenomena to which this theory may be applied Quantum mechanics also termed wave mechanics generalizes and supersedes the older classical mechanics and Maxwell s electromagnetic theory Atomic and subatomic phenomena provide the most striking evidence for the correctness of quantum mechanics and also best illustrate the differences between quantum mechanics and the older classical physical theories Quantum mechanics is needed to explain many properties of bulk matter for instance the temperature dependence of the specific heats of solids These along with numerous other applications are more fully discussed in the articles and bibliography to which this article refers

The formalism of quantum mechanics is not the same in all domains of applicability In approximate order of increasing conceptual difficulty mathematical complexity and likelihood of future fundamental revision these domains are (1) nonrelativistic quantum mechanics applicable to systems in which particles are neither created nor destroyed and in which the particles are moving slowly compared to the velocity of light

$$(c \simeq 3 \times 10^{10} \text{ cm/sec})$$

Here a particle is defined as a material entity having mass whose internal structure either does not change or is irrelevant to the description of the system (2) relativistic quantum mechanics applicable in practice to a single relativistic particle (one whose speed equals or nearly equals c) here the particle may have zero rest mass in which event its speed must equal c (3) quantum field theory applicable to systems in which particle creation and destruction can occur the particles may have zero or nonzero rest mass

This article is concerned mainly with nonrelativistic quantum mechanics which apparently applies to all atomic and molecular phenomena with the exception of the finer details of atomic spectra (see ATOMIC STRUCTURE AND SPECTRA) Nonrelativistic quantum mech[illegible]
in [illegible]
ing [illegible]
kin [illegible] 10 ev (1 ev = 1 electron volt = 1.6×10^{-12} erg is the energy gained by an electron in traversing a potential difference of 1 volt) Many quantum mechanical predictions are not as quantitatively accurate for nuclei as for atomic and molecular systems however because nuclear forces are not yet accurately known

For the formal mathematical structure of nonrelativistic quantum mechanics *see* QUANTUM THEORY NONRELATIVISTIC That article provides justification for many of the assertions made under the subheadings immediately following wherein are described the novel (from the standpoint of classical physics) features of nonrelativistic quantum mechanics Some of these features are retained others modified in the more complicated domains of relativistic quantum mechanics and quantum field theory *See* QUANTUM ELECTRODYNAMICS QUANTUM FIELD THEORY QUANTUM THEORY RELATIVISTIC SYMMETRY LAWS (PHYSICS)

Planck s constant The quantity 6.61×10^{-27} erg-sec first introduced into physical theory by Max Planck in 1901 is a basic ingredient of the formalism of quantum mechanics Most of the fundamental quantum mechanical relations for example Schrodinger s equation and Heisenberg s

Nuclear forces, mesons. The fields which manifest themselves only at the atomic level may be understood in terms of an analogy The electromagnetic field is the agency of the electrical forces which account for the properties of atoms Similarly, the structure of the nucleus reveals the existence of a nuclear force which accounts for the interaction between the particles found in the nucleus The nuclear force is strong enough to balance the electrical repulsive forces due to the charged protons found in the nucleus, it is of short range, effective only when the nuclear particles are within a certain radius This is in contrast to the infinite range of the Coulomb electrical forces One is led to make the analogy, with appropriate modifications, between the electrical and nuclear forces and to suppose that a field accounts for the nuclear force Because such a field has its own degrees of freedom, it is capable of an independent existence just as the electromagnetic field is capable of carrying energy in the form of electromagnetic waves When such a field is "quantized" the field system exhibits particle properties

One can estimate the mass of the particles by describing the force field which surrounds a nuclear particle in terms of the emission and absorption of the quanta of the field Thus, one couples the appropriate nuclear coordinate linearly to the field ϕ and hence to the emission and absorption operators $\phi^{( )}$ If the mass of such quanta is m, when the nucleon (nuclear particle) radiates such a particle the energy cost is $\sim mc^2$ According to the uncertainty principle, if the energy of a system is to be observed to the accuracy ΔE, the observation must be made over an interval Δt longer than $\hbar/\Delta E$ (*see* UNCERTAINTY PRINCIPLE) Thus, an energy fluctuation of a nucleon of magnitude mc^2 can take place over an interval $\Delta t < \hbar/mc^2$ without violation of the energy principle In this time interval the field quantum can travel a distance not exceeding $c\Delta t < \hbar/mc$ before it must be reabsorbed by the nucleon Hence, the measure of the range of the force field which surrounds a nucleon is the Compton wavelength, $\hbar/mc$, and if this distance is set equal to the known range of the nuclear force one obtains an estimation of the mass of the quantum of the field In this way H Yukawa was led to predict the existence of a particle, the meson, which is the quantum of the nuclear force field and which has a mass ($270m_e$) intermediate between that of the electron ($m_e = 9.1 \times 10^{-28}$ g) and the nucleon ($1840m_e$)

The Yukawa meson has since been identified as the π-meson, of which there are three varieties, two with electrical charges equal in magnitude but opposite in sign, and one electrically neutral The free π meson field proves not to be stable, that is, the charged varieties of mesons ($\pi^{\pm}$) decay in a time of the order of 10^{-8} sec into quanta of other fields while the neutral variety (π^0) transforms itself into two quanta of the electromagnetic field (γ rays) in a time less than 10^{-15} sec Hence as a field with an independent existence, its presence is too ephemeral to have been observed classically, but, more important, as a force field its range is too short to have been detected macroscopically *See* MESON NUCLEAR STRUCTURE

Particle fields. In addition to the "force fields" which act between the particles and which give rise to the existence of other particles one may ask if particles such as electrons and neutrons are the quanta of fields The existence of the pair production of electrons and positrons, neutrons and antineutrons, etc, illustrates that these particles may also be viewed as the quanta of fields. Indeed, one may also regard these particle pairs as producing force fields between the quanta of the other fields Thus, the quanta of the electromagnetic field or photons, are coupled to pairs of the electron positron field and this coupling leads to a force between photons, or the scattering of light by light There are two general types of field corresponding to the two general classes of statistics obeyed by elementary particles Particles which follow the exclusion principle (these have Fermi Dirac statistics) are characterized by fields which carry an internal angular momentum or spin which is half integer while the particles which have Bose Einstein statistics are of integer spin One of the greatest successes of the quantum theory of fields together with the relativity principle was the prediction of the observed connection between spin and statistics *See* QUANTUM STATISTICS

Solution of field equations The general problem in the quantum theory of fields may be stated as follows Sets of fields are introduced to describe the various types of observed elementary particles The interactions among the particles are produced by coupling the fields and the attempt is made to deduce the types of coupling which should be present from the experimentally observed interactions among the particles

Perturbations The difficulties associated with this program are formidable The existence of a field description alone implies that any coupling of the fields leads to an enormous class of distinct types of interactions Thus there is an infinite class of states for the fields and because of practical limitations coupling must be deduced by studying only a limited number experimentally Thus one is confronted with a gigantic puzzle with only a few pieces of information Moreover great mathematical difficulties are encountered in trying to solve the coupled field equations The standard methods of quantum mechanical perturbation theory are not adequate unless the coupling is rather weak However, the known couplings in the case of the subnuclear interactions are quite strong

At present no adequate methods exist for the solution of the field equations when the coupling is strong Further even in the case of the weak couplings problems arise in the straightforward application of perturbation methods These

tron) is proportional to the square root of the inci dent light intensity and is not directly related to nor limited by the incident light frequency f (see ELECTROMAGNETIC RADIATION MAXWELL'S EQUATIONS) The photoelectric effect is interpreted readily on the assumptions that the energy in the light beam is carried in quanta of energy $E = hf$, that emission of an electron results from absorption of a single quantum (a single photon), and that absorption of half or any fraction of a quantum is not possible because the photons act like discrete indivisible entities For additional information on the photoelectric effect see PHOTOEMISSION The particle properties of electromagnetic waves also are demonstrated in the Compton effect, wherein the wavelengths of x rays are lengthened by scattering from free electrons The change in wavelength is predicted quantitatively, assuming the scattering results from elastic collisions between photons and electrons, and using Eqs (2a) and (2b) for the photon momentum and energy The diffraction of x rays by crystals was regarded in the prequantum era as conclusive proof that x rays are waves and not "corpuscles" See COMPTON EFFECT X RAY DIFFRACTION

Interference and diffraction Wave propagation is distinguished from particle propagation by the phenomena of interference and diffraction It is a general result of wave theories that interference and diffraction effects largely are confined to an angle (relative to the incident beam) which in radi

u iracting the wave or the distance between two interfering scattering centers (see DIFFRACTION, INTERFERENCE OF WAVES) This fact and the magnitudes of λ inferred from Eq (2a) are sufficient to explain why wave effects are not observed in the propagation of ordinary macroscopic bodies but can be observed in the propagation of electrons neutrons and light atoms or molecules For example for a mass of 1 g moving at 1 cm/sec $\lambda = 6.6 \times 10^{-27}$ cm But for a neutron or hydrogen atom (using $p = Mv = \sqrt{2ME}$ where $M = 1.66 \times 10^{-24}$ g) moving at velocity corresponding to room temperature (300°K) $E = 3/2\ kT$ (k = Boltzmann's constant = 1.38×10^{-16} erg deg^{-1}) and λ turns out to equal 1.45×10^{-8} cm For an electron with an energy of 100 ev $\lambda = 1.22 \times 10^{-8}$ cm For a proton with an energy 10^6 ev = 1 Mev $\lambda = 2.88 \times 10^{-12}$ cm These numerical results and the discussion of this paragraph also explain the ability of crystals wherein interatomic spacings are about 10^{-8} cm to give a good demonstration of electron and molecular diffraction suggest the need for quantum mechanics to "understand" atomic systems wherein atomic dimensions are $\sim 10^{-8}$ cm and electron energies are ~ 10 ev suggest the need for quantum mechanics to 'understand" atomic nuclei wherein nuclear dimensions are $\sim 10^{-13}$ cm and neutron or proton energies are ~ 10 Mev and explain why quantum effects are more readily observed in H_2 and He than in heavier gases and at low temperatures rather than high

Relationship to uncertainty principle Wave particle duality is intimately connected with the uncertainty principle in that the uncertainty inequalities can be derived from analyses of specific experiments For a nonrelativistic particle the connection can be seen from the following argument which contains the basic elements of the rigorous formal treatment As explained later, the probability that the x coordinate of the particle will be in the interval x to $x + dx$ is $|\psi(x)|^2\, dx$, where $\psi(x)$ is called the wave function Suppose measurement has ascertained that the particle lies in an interval of width Δx centered at $x = 0$, that is, measurement has determined that the dependence of $\psi(x)$ on x is approximately as shown in Fig. 1 Because of wave-particle duality the wave-packet of Fig 1 can be looked upon as a superposition of waves Since $\psi(x)$ rises from and falls to a very small value in an interval Δx, the packet must contain waves whose half wavelengths are as small as Δx, furthermore it can be proved that because $\psi(x)$ does not change sign the packet must contain waves of very long wavelength Thus, in the packet, the wavelengths λ run from about $2\Delta x$ to ∞, the reciprocal wavelengths λ^{-1} run from about 0 to $1/(2\Delta x)$ But from Eq (2a), $\Delta p_x = h\Delta\lambda^{-1}$ Hence $\Delta p_x\, \Delta x$ cannot be less than about $h/2$ Considering the simplicity of the argument, this result is close enough to Eq (1a) Similarly, Eq (1d) can be understood from Eq (2b) and from the fact that to transmit information in a time Δt, that is to turn a measuring instrument on and off in an interval Δt it is necessary to use frequencies higher than about $(2\Delta t)^{-1}$ Using Eq (1a), the discussion of this paragraph suggests that in atoms whose dimensions Δx are about 10^{-8} cm one must expect to find electrons with energies

$$E = p^2/2m = h^2/2m(\Delta x)^2 = 3.8 \text{ ev}$$

which is of the order of magnitude observed Similarly in atomic nuclei whose dimensions Δx are about 10^{-13} cm one must expect to find neutrons or protons with energies about 20 Mev, again, of the order of magnitude observed On the other hand if atomic nuclei contained electrons the

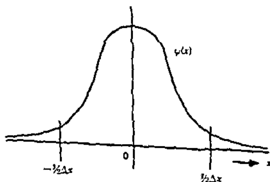


Fig 1 Plot of $\psi(x)$ versus x

uncertainty principle explicitly involve Planck's constant as do many of the well verified consequences of quantum mechanics for example, the formula for the energy levels of atomic hydrogen Planck's constant plays no role in the classical theories Planck's constant commonly is denoted by the letter h the notation $\hbar = h/2\pi$ also is standard

Uncertainty principle In classical physics the observables characterizing a given system are assumed to be simultaneously measurable (in principle) with arbitrarily small error For instance it is thought possible to observe the initial position and velocity of a particle and therewith using Newton's laws to predict exactly its future path in any assigned force field According to the uncertainty principle (W Heisenberg 1927) accurate measurement of an observable quantity necessarily produces uncertainties in one's knowledge of the values of other observables In particular for a single particle

$$\Delta x \, \Delta p_x \gtrsim \hbar \qquad (1a)$$

where Δx represents the uncertainty (error) in the location of the x coordinate of the particle at any instant and Δp_x is the simultaneous uncertainty in the x component of the particle momentum Equation (1a) asserts that under the best circumstances the product $\Delta x \, \Delta p_x$ of the uncertainties cannot be less than about 10^{-27} erg sec of course with poor measurements the product can be much greater than $\hbar$ On the other hand there is no restriction on the simultaneous determination of position along x and momentum along y that is the product $\Delta x \, \Delta p_y$ may equal zero Other typical uncertainty inequalities for a single particle are

$$\Delta\phi \, \Delta l_z \gtrsim \hbar \qquad (1b)$$

$$\Delta x \, \Delta E \gtrsim \frac{\hbar}{m} p_x \qquad (1c)$$

In Eq (1b) the particle location is specified in spherical coordinates with polar axis along z $\Delta\phi$ is the uncertainty in azimuth angle Δl_z is the uncertainty in the z component of the orbital angular momentum In Eq (1c) ΔE is the uncertainty in energy, m is the particle mass The uncertainty relation

$$\Delta t \, \Delta E \gtrsim \hbar \qquad (1d)$$

is derived and interpreted somewhat differently than Eqs (1a) to (1c) it asserts that for any system an energy measurement with error ΔE must be performed in a time not less than $\Delta t \sim \hbar/\Delta E$ If a system endures for only Δt sec any measurement of its energy must be uncertain by at least $\Delta E \sim \hbar/\Delta t$ ergs

Because the numerical value of $\hbar$ is so small and since $\Delta p_x = m\Delta v_x$ v denoting velocity the restrictions implied by Eqs (1a) to (1c) are utterly inconsequential for macroscopic systems wherein masses are of the order of grams For an electron however whose mass is 9.1×10^{-28} g $\hbar/m \sim 1$ cm²/sec and the uncertainty principle cannot be ignored Similarly Eq (1d) is unimportant for macroscopic systems wherein energies are of the order of ergs but is significant for atomic systems where $\Delta E = \hbar/\Delta t$ need not be negligible compared to the actual energy E See UNCERTAINTY PRINCIPLE

Wave particle duality It is natural to identify such fundamental constituents of matter as protons and electrons with the mass points or particles of classical mechanics According to quantum mechanics however these particles in fact all material systems necessarily have wavelike properties Conversely the propagation of light, which by Maxwell's electromagnetic theory is understood to be a wave phenomenon is associated in quantum mechanics with massless energetic and momentum transporting particles called photons (see PHOTON) The quantum mechanical synthesis of wave and particle concepts is embodied in the de Broglie relations

$$\lambda = h/p \qquad (2a)$$

$$f = E/h \qquad (2b)$$

These give the wavelength λ and wave frequency f associated with a free particle (a particle moving freely under no forces) whose momentum is p and energy is E the same relations give the photon momentum p and energy E associated with an electromagnetic wave in free space (that is in a vacuum) whose wavelength is λ and frequency is f

The wave properties of matter have been demonstrated conclusively for beams of electrons of neutrons of atoms (hydrogen H and helium He) and of molecules (H_2) When incident upon crystals these beams are reflected into certain directions forming diffraction patterns Diffraction patterns are difficult to explain on a particle picture they are readily understood on a wave picture in which wavelets scattered from regularly spaced atoms in the crystal lattice interfere constructively along certain directions only Moreover the wavelengths of these matter waves as inferred from the diffraction patterns agree with the values computed from Eq (2a) as first demonstrated by C J Davisson and L H Germer in 1927 See ELECTRON DIFFRACTION NEUTRON DIFFRACTION

Photoelectric effect Compton effect The particle properties of light waves are observed in the photoelectric effect When light of frequency f causes electrons to be emitted from a surface all the electrons have very nearly the same maximum kinetic energy the maximum kinetic energy is independent of the light intensity the number of electrons emitted in unit time is proportional to the light intensity as f is varied the maximum electron kinetic energy W varies linearly with f in fact $W = hf - C$ C being a constant characteristic of the emitting material These observations are difficult to understand on the wave picture wherein the magnitude of the electric field vector (which presumably exerts the force which ejects the elec-

han actually is obtainable See SUPERPOSITION PRINCIPLE OF

The limitations on the experiments are imposed by the requirements of the uncertainty principle Eqs (1) and wave particle duality Eqs (2) To illustrate the analysis suppose indicators (symbolized by circles) free to move in the vertical y direction are placed immediately behind the slits S_1 and S_2 in Fig 2 the recoil of an indicator signifies a collision with a photon and places the photon at the indicator In order that the recoil establish the slit through which the photon has passed the uncertainty in the vertical location of each indicator must be much less than $d/2$ According to Eq (1a) this means that the vertical momentum p_y of each indicator will be uncertain by an amount $\Delta p_y >> 2\hbar/d$ There can be no assurance that the indicator has recoiled unless δp_y, the momentum transferred from photon to indicator equals or exceeds Δp_y When the slits are narrow and $X >> d$ a momentum transfer δp_y deflects the photon through an angle $\theta \sim \delta p_y/p$ and changes the point I where the photon strikes the screen by an amount ΔY $\theta X \approx X\, \delta p_y/p$ with $p = h/\lambda$ Eq (2a) It follows that when there are observable recoils localizing the photon at one of the slits then $\Delta Y >> \lambda X/\pi d$ that is $\Delta Y >> Y/\pi$ where $Y/2 = \lambda X/2d$ is the distance between adjacent maxima and minima of the interference pattern Thus determining the slit through which the photon passes necessarily gives the photon an uncontrollable random vertical deflection (on the screen Q_2) which is very much larger than the distance between adjacent maxima and minima. The photons now spread with uniform intensity over vertical distances ΔY which are large compared to the original widths of the light and dark interference bands in other words the interference pattern is destroyed

Quantization In classical physics the possible numerical values of each observable, meaning the possible results of exact measurement of the observable generally form a continuous set. For example the x coordinate of the position of a particle may have any value between $-\infty$ and $+\infty$ similarly p the x component of momentum may have any value between $-\infty$ and $+\infty$ the kinetic energy T of a particle may have any value between 0 and $+\infty$ the total energy (kinetic plus potential) of an electron in the field of a proton may have any value between $-\infty$ and $+\infty$ the orbital angular momentum vector $\mathbf{l}$ of a particle moving in a central field for example an electron in a hydrogen atom, may have any magnitude between 0 and ∞ if the magnitude of $\mathbf{l}$ is known to be l then since the plane of the orbit may be arbitrarily oriented the z component of $\mathbf{l}$ may have any value between $-l$ and $+l$ In quantum mechanics the possible numerical values of an observable need not form a continuous set, however For some observables the possible results of exact measurement form a discrete set for other observables the possible numerical values are partly discrete partly continuous for example the total energy of an electron in the field of a proton may have any positive value between 0 and $+\infty$ but may have only a discrete set of negative values namely -13.6 $-13.6/4$ $-13.6/9$ $-13.6/16$ ev Such observables are said to be quantized often there are simple quantization rules determining the quantum numbers which specify the allowable discrete values (see QUANTUM NUMBERS) For example $\mathbf{L}^2$ the square of the orbital angular momentum of a particle must equal $l(l+1)\hbar^2$ where l is zero or a positive integer, the only allowed values of $\mathbf{J}$ the square of the total angular momentum (orbital angular momentum plus intrinsic angular momentum or spin) are $J^2 = j(j+1)\hbar^2$ where $j = 0$ 1/2 1 3/2, 2 that is $j\hbar$ is an integral or half integral multiple of $h/2\pi$ If the magnitude of the total angular momentum is given by some one of these values $j\hbar = 3\hbar/2$ for example then the only allowed values of the z component J_z of $\mathbf{J}$ are $-3\hbar/2$ $-\hbar/2$ $\hbar/2$ $3\hbar/2$ that is for a given j the allowed values of $J_z = -j$ $-j+1$ $-j+2$ ...

... same allowed values in quantum mechanics as they do classically In the formal theory each observable is a linear operator whose eigenvalues (characteristic values) are the allowed values of that observable The set of eigenvalues is termed the spectrum of the operator which spectrum may be d screte ...

... illustrated in the photoelectric and Compton effects described earlier

Stern Gerlach experiment The quantization of angular momentum is strikingly exhibited in the Stern Gerlach experiment wherein a beam of for example hydrogen atoms moving through a region of space containing an inhomogeneous magnetic field breaks up into two separate beams Classically the force on a hydrogen atom in an inhomogeneous magnetic field depends on the angle between the magnetic field and the plane of the electron s orbit Thus classically the original beam in which all orientations of the plane of the orbit are possible, is expected to spread out or defocus in the inhomogeneous magnet ... is not ...

Fo ...

qu ... tion that the square of the total angular momentum of atomic hydrogen is $J^2 = 3\hbar^2/4 = \frac{1}{2}(\frac{1}{2}+1)\hbar^2$ and therefore that the vector $\mathbf{J}$ (which classically always is perpendicular to the orbit plane) must be either parallel or antiparallel to the applied magnetic field (along z) for these two directions of $\mathbf{J}$ and only for these two the value of J_z equals one of the permitted values $J_z = -\hbar/2$ or $J_z = \hbar/2$. The Stern Gerlach experiment also can be performed with beams of

electron energies would be as high as 200 Mev In computing this value it is necessary to use the relativistic relation $E = \sqrt{m^2c^4 + c^2p^2}$ connecting E and p (*see* RELATIVISTIC MECHANICS) These energies are much too high to be explained by electrostatic forces between electrons and protons at separations of about 10^{-13} cm Thus the uncertainty principle leads to the inference that electrons are not contained in atomic nuclei which inference is in accord with present (1960) theories of nuclear structure and β decay *See* NUCLEAR STRUCTURE RADIOACTIVITY

Complementarity Wave particle duality and the uncertainty principle are thought to be examples of the more profound principle of complementarity first enunciated by Niels Bohr (1928) According to the principle of complementarity nature has complementary aspects an experiment which illuminates one of these aspects necessarily simultaneously obscures the complementary aspect To put it differently each experiment or sequence of experiments yields only a limited amount of information about the system under investigation as this information is gained other equally interesting information (which could have been obtained from another sequence of experiments) is lost Of course the experimenter does not forget the results of previous experiments but at any instant only a limited amount of information—the information gained from the most recent experiment—is usable for predicting the future course of the system

The well known double slit experiment provides a good illustration of the principle of complementarity Light from a monochromatic point source P (Fig 2) is diffracted by the two slits S_1 and S_2 in the screen Q_1 On the screen Q_2 an interference pattern of alternating bright and dark bands is formed in the region D_1D where the two diffraction patterns (from the slits S_1 and S_2) overlap Assuming that P is equidistant from the slits and also that the slits are very narrow compared to the distances PS_1 or S_1O it follows that interference maxima (bright bands) are observed whenever $S_2O - S_1O$ equals $n\lambda$ λ being the wavelength of the light and n an integer when $S_2O - S_1O = (n + \frac{1}{2})\lambda$ interference minima (dark bands) are observed In moving from any maximum to an adjacent maximum the path difference $S_2O - S_1O$ changes by precisely one wavelength Consequently measurement of the distance Y between successive maxima and knowledge of S_1S_2 and of the distance X between Q_1 and Q_2 yields λ via the formula $\lambda = Yd/X$ (valid when $S_1S_2 = d$ is much smaller than X) Evidently the double-slit experiment is understandable in terms of and provides information concerning the wave properties of light

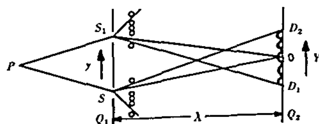


Fig 2 Double-slit experiment

The double slit experiment yields no information concerning the particle properties of light in fact introducing the particle picture leads only to conceptual difficulties These difficulties appear with the recognition that reducing the source intensity does not modify the interference pattern after a sufficiently long exposure a photographic film at Q_2 will show exactly the same interference pattern D_1D_2 as is observed by the eye using a more intense source Since it is possible to make the source intensity so low that two photons almost never will be emitted during the very small time required for light to travel from P to Q_2 via either of the slits it is necessary to conclude that the interference pattern is produced by independent individual photons and not by interference between two or more different photons On the other hand the interference pattern is destroyed when either of the slits is closed Thus the question arises How can a stream of independent photons each of which presumably passes through only one of the slits and half of which on the average pass through S_1 produce an interference pattern that is destroyed by closing one of the slits? Or to put it differently how can closing or opening a slit through which a photon does not pass affect the likelihood of that photon reaching any particular point on Q_2?

The principle of complementarity meets these difficulties with the assertion that the possibility of demonstrating that the photons have well defined trajectories through one or the other slit (a particlelike property) is complementary to the possibility of demonstrating the wavelike property of interference In the double slit experimental setup which has been described until the photon is localized at Q_2 (by the visible evidence that a chemical effect has occurred in a photographic film) it is not possible to locate the photon at any particular point in space nor is it legitimate to insist that the photon must have passed through only one of the slits Moreover according to the principle of complementarity modifying the experimental setup so as to localize the photon at one of the slits and thereby to determine through which slit the photon passes necessarily destroys the interference pattern This last assertion is supported by analysis of various photon detection schemes recognizing that the proposed experiments perforce are entirely *Gedanken* (in the mind) an actual measurement of the slit through which the photon passes demands extreme precision and to this date (1960) has not been attempted It is concluded that quantum mechanics involves no inconsistencies or paradoxes From the standpoint of the complementarity principle the questions of the preceding paragraph and other similar difficulties rest always on the specious assumption of more information

$$P_\alpha(\beta) = \left| \int_{-\infty}^{\infty} dx\, v^*(x,\beta) u(x\,\alpha) \right|^2 \qquad (6)$$

where $u(x\,\alpha)$ is the eigenfunction corresponding to $A = \alpha$ $v(x\,\beta)$ is the eigenfunction corresponding to $B = \beta$ and the asterisk denotes the complex conjugate Since measurement of A in the state $A = \alpha$ must yield the result $A = \alpha$ it is necessary that the states $u(x\,\alpha)$ satisfy the normalizing and orthogonalizing relation

$$\left| \int_{-\infty}^{\infty} dx\, u^*(x,\alpha') u(x,\alpha) \right|^2 = \delta_{\alpha\alpha'} \qquad (7)$$

where $\delta_{\alpha\alpha'} = 0$ when $\alpha \neq \alpha'$ (eigenfunctions orthogonal) and $\delta_{\alpha\alpha} = 1$ when $\alpha = \alpha'$ (eigenfunctions normalized) the eigenfunctions $v(x\,\beta)$ are similarly orthonormal The integral in Eq (6) is called the projection of $u(x\,\alpha)$ on $v(x\,\beta)$ The projection of the eigenfunction corresponding to $A = \alpha$ on the eigenfunctions of the x operator is $u(x,\alpha)$ and $|u(x,\alpha)|^2\, dx$ is the probability that the system known to be in the eigenstate $A = \alpha$ will be found in the interval x to $x + dx$

The formalism just described embodies the essential feature that each measurement on an individual system as it develops new information necessarily loses or makes untrue some information gained in the past in fact this formalism leads to a rigorous derivation of the uncertainty relations (1) For example suppose B is measured with the system in the state $A = \alpha$ and it is found that $B = \beta$ exactly The act of measurement necessarily and unavoidably disturbs the system with the result that after the measurement the system is in the eigenstate $v(\beta,x)$ After the measurement therefore it no longer is certain that $A = \alpha$ rather the probability of finding $A = \alpha$ is $P_\beta(\alpha)$ which by Eq (6) equals $P_\alpha(\beta)$ Thus after starting with $A = \alpha$ and then measuring B β it is possible to find that A equals $\alpha' \neq \alpha$ Of course as stressed previously the e considerations are unimportant for macroscopic systems where the limitations imposed by the uncertainty principle are inconsequential Even if all observations were extremely accurate by usual standards when the momentum p of a 1 g mass at x is measured the position x' immediately thereafter should be indistinguishable from x Were x' and x distinguishable the particle position seemingly would have changed discontinuously from x to x' contrary to all classical (macroscopic) experience

This formalism yields predictions in excellent agreement with observation furthermore it can be seen that the formalism is internally consistent Consequently the fol
the t
it is
ber o identical systems, the result of a measurement on a single (microscopic) system generally is unpredictable and largely a matter of chance Nonetheless some ob
have t
of
tory to put the problem in simplest terms consider a gram of radium containing approximately 10^{21} atoms According to generally accepted theory it is not possible to predict when any one atom will decay but it is possible to predict very accurately the average number of atoms decaying every second The objectors to this doctrine feel that it must be possible to predict the subsequent history of every individual atom failure to do so represents not an inherent indeterminism in nature but rather a lack of obtainable information—and therefore a lack of understanding—concerning the mechanism of radioactive decay To mention but one possible alternative nonrelativistic quantum theory can be reinterpreted in terms of hidden variables which in principle determine the precise behavior of an individual system but whose values are not ascertained in measurements of the type which now can be carried out This alternative has not led to new predictions however and contains some unappealingly *ad hoc* features See CAUSALITY PROBABILITY IN PHYSICS

Wave function When the system is known to be in the eigenstate corresponding to $A = \alpha$ the

gives the prob-ability of measuring $B = \beta$ The wave function $\psi(x)$ may be known exactly in other words the state of the system may be known as exactly as possible (within the limitations of uncertainty and complementarity) even though $\psi(x)$ is not the eigenfunction of a known operator This circumstance arises because the wave function obeys Schrodinger s wave equation Knowing the value of $\psi(x)$ at time $t = 0$ the wave equation completely determines $\psi(x)$ at all future times In general however if $\psi(x\,0) = u(x\,\alpha)$ that is if $\psi(x\,t)$ is an eigenfunction of A at $t = 0$ then $\psi(x\,t)$ will not be an eigenfunction of A at later times $t > 0$ For example suppose at $t = 0$ a free particle (a particle moving under no forces) is known to be in an eigenstate for which the uncertainty in x is $(\Delta x)_0$ $(\Delta x)_0$ is approximately the x interval within which $|\psi(x\,0)|^2$ is not negligibly small Suppose further that at $t = 0$ the product of the uncertainties in position and momentum is as small as possible $(\Delta x)_0(\Delta p_x)_0 \cong \hbar$ compare Eq (1a) Then it can be proved that $(\Delta x)_t$ the uncertainty in x at time t satisfies

$$(\Delta x)_t = \left[(\Delta x)_0{}^2 + \frac{t^2}{m^2} (\Delta p_x)_0{}^2 \right]^{1/2} \qquad (8)$$

where m is the particle mass Equation (8) is readily interpreted The root mean square spread at time t results from $(\Delta x)_0$ and from an uncertainty in the distance the particle has traveled the latter uncertainty is $t(\Delta v_x)_0 = t(\Delta p_x)_0/m$ If $(\Delta x)_0 = 0$ meaning $\psi(x\,0)$ is an eigenfunction of the x operator $(\Delta x)_t$ is infinite showing that $\psi(x,t)$ cannot be an eigenfunction of the x operator When the particle is free the projections of the wave function on the eigenfunctions of the mo-

other atoms and molecules as well as with neutron beams In all cases the observations agree with quantum mechanical expectation Refinements of the experiment yield accurate measurements of the parameters characterizing the magnetic interactions of atoms molecules, and atomic nuclei *See* MAGNETIC RESONANCE, MOLECULAR BEAMS, NUCLEAR MOMENTS

Atomic spectra Spectroscopy, especially the study of atomic spectra probably provides the most detailed quantitative confirmation of quantization Granting that the energy in a light wave is quantized it follows from conservation of energy and Eq (2*b*) that when an atom emits a photon of frequency f, its initial energy E_i and final energy E_f after emission are related by the equation

$$E_i - E_f = hf \tag{3a}$$

Similarly, when a photon of frequency f is absorbed the initial energy E'_i, and final energy E'_f after absorption satisfy

$$E'_f - E'_i = hf \tag{3b}$$

In quantum mechanics as in classical mechanics an electron remains in the vicinity of a proton, that is it is "bound" to the proton when and only when the magnitude of its kinetic energy T is less than the magnitude of its negative potential energy V (V is set equal to zero at infinite separation) Thus, the total energy of a stable hydrogen atom necessarily is negative, in other words the only allowed energies of atomic hydrogen are $-R/n^2$, where n is an integer and $R = -13.6$ ev Consequently the radiation emitted by atomic hydrogen must consist of a discrete set of frequencies, or lines obeying the relation from Eq (3*a*),

$$f = \frac{R}{h}\left(\frac{1}{n^2} - \frac{1}{m^2}\right) \tag{4}$$

where m and n are integers $m > n$ This simple argument provides a convincing explanation of the observation that atomic hydrogen has a line spectrum and not a continuous spectrum, in other words it radiates discrete frequencies rather than a continuous band of frequencies The observed lines very accurately satisfy Eq (4), moreover, except for small relativistic and quantum field theory effects the nonrelativistic Schrödinger equation accurately predicts not merely the frequencies of the lines but also their relative intensities and widths Because on the average the levels E_i and E_f of Eq (3*a*) endure only for a limited time, their energies are uncertain by an amount ΔE, as explained in the earlier discussion of the uncertainty principle, therefore, the observed lines have a width not less than $\Delta f \cong h^{-1}(\Delta E_i + \Delta E_f)$ according to Eq (3*a*) Heavier atoms have more complicated line spectra than hydrogen and the frequencies they emit cannot be described by formulas as simple as Eq (4) For these atoms the *agreement* between experiment and the predictions of the nonrelativistic Schrödinger equation though always very good is not as precise as for hydrogen All the evidence indicates, however, that the discrepancies arise because the Schrodinger equation cannot be solved exactly in many electron atoms, so that the theoretical predictions necessarily are approximate If approximations were not necessary (for example if perfect computing machines were available), there is every reason to think that the predictions of the Schrödinger equation would be exactly correct, except for the aforementioned small relativistic and field theory effects

Probability considerations The uncertainty and complementarity principles which limit the experimenter's ability to describe a physical system, must limit equally the experimenter's ability to predict the results of measurement on that system. Suppose, for instance, that a very careful measurement determines that the x coordinate of a particle is precisely $x = x_0$ This is permissible in nonrelativistic quantum mechanics Then formally the particle is known to be in the eigenstate corresponding to the eigenvalue $x = x_0$ of the x operator Under these circumstances an immediate repetition of the position measurement again will indicate that the particle lies at $x = x_0$, if the particle is moving in a one-dimensional force field described by the potential $V(x)$, the particle's potential energy will be exactly $V(x_0)$ Knowing that the particle lies at $x = x_0$ makes the momentum p_x of the particle completely uncertain however according to Eq (1*a*) A measurement of p_x immediately after the particle is located at $x = x_0$ could yield any value of p_x from $-\infty$ to $+\infty$, a measurement of $T = p_x^2/2m$ could yield any value from 0 to $+\infty$ and in fact the average or expectation value of T in these circumstances would be infinite

More generally suppose the system *is known to* be in the eigenstate corresponding to the eigenvalue α of the observable A Then for any observable B, which is to some extent complementary to A, that is for which an uncertainty relation of the form of Eqs (1) limits the accuracy with which A and B can simultaneously be measured it is not possible to predict which of the many possible values $B = \beta$ will be observed However it is possible to predict the relative probabilities $P_\alpha(\beta)$ of immediately thereafter finding the observable B equal to β that is of finding the system in the eigenstate corresponding to the eigenvalue $B = \beta$ If the system is prepared in the eigenstate α of A a great many times and each time the observable B is measured immediately thereafter the average of these observed values of B will equal the expectation value of B defined as

$$\langle B\rangle = \Sigma \beta P_\alpha(\beta) \tag{5}$$

summed over all eigenvalues of B when the spectrum is continuous the summation sign is replaced by an integral To the eigenvalues correspond eigenfunctions in terms of which $P_\alpha(\beta)$ can be computed In particular, when α is a discrete eigenvalue of A, and the operators depend only on x and p_x, the probability $P_\alpha(\beta)$ is postulated to be

$$P_\alpha(\beta) = \left| \int_{-\infty}^{\infty} dx\, v^*(x,\beta) u(x,\alpha) \right|^2 \qquad (6)$$

where $u(x,\alpha)$ is the eigenfunction corresponding to $A = \alpha$ $v(x\,\beta)$ is the eigenfunction corresponding to $B = \beta$ and the asterisk denotes the complex conjugate Since measurement of A in the state $A = \alpha$ must yield the result $A = \alpha$ it is necessary that the states $u(x\,\alpha)$ satisfy the normalizing and orthogonalizing relation

$$\left| \int_{-\infty}^{\infty} dx\, u^*(x\,\alpha') u(x\,\alpha) \right|^2 = \delta_{\alpha\alpha'} \qquad (7)$$

where $\delta_{\alpha\alpha'} = 0$ when $\alpha \neq \alpha'$ (eigenfunctions orthogonal) and $\delta_{\alpha\alpha'} = 1$ when $\alpha = \alpha'$ (eigenfunctions normalized) the eigenfunctions $v(x,\beta)$ are similarly orthonormal The integral in Eq (6) is called the projection of $u(x\,\alpha)$ on $v(x,\beta)$ The projection of the eigenfunction corresponding to $A = \alpha$ on the eigenfunctions of the x operator is $u(x,\alpha)$ and $|u(x\,\alpha)|^2\,dx$ is the probability that the system known to be in the eigenstate $A = \alpha$, will be found in the interval x to $x + dx$

The formalism just described embodies the essential feature that each measurement on an individual system as it develops new information necessarily loses or makes untrue some information gained in the past in fact this formalism leads to a rigorous derivation of the uncertainty relations (1) For example suppose B is measured with the system in the state $A = \alpha$ and it is found that $B = \beta$ exactly The act of measurement necessarily and unavoidably disturbs the system with the result that after the measurement the system is in the eigenstate $v(\beta,x)$ After the measurement, therefore it no longer is certain that $A = \alpha$ rather the probability of finding $A = \alpha$ is $P_\beta(\alpha)$ which by Eq (6) equals $P_\alpha(\beta)$ Thus after starting with $A = \alpha$ and then measuring $B = \beta$ it is possible to find that A equals $\alpha' \neq \alpha$ Of course as stressed previously these considerations are unimportant for macroscopic systems where the limitations imposed by the uncertainty principle are inconsequential Even if all observations were extremely accurate by usual standards, when the momentum p of a 1 g mass at x is measured the position x' immediately thereafter should be indistinguishable from x Were x' and x distinguishable the particle position seemingly would have changed discontinuously from x to x' contrary to all classical (macroscopic) experience

This formalism yields predictions in excellent agreement with observation furthermore it can be seen that the formalism is internally consistent Consequently the f u
the formalis
it is possible
ber of obse n identical systems, the result of a measurement on a single (microscopic) system generally is unpredictable and largely a matter of chance Nonetheless, some ob
have ral
of na
defici
, to put the problem in simplest terms consider a gram of radium containing approximately $10^{?}$ atoms According to generally accepted theory it is not possible to predict when any one atom will decay but it is possible to predict very accurately the average number of atoms decaying every second The objectors to this doctrine feel that it must be possible to predict the subsequent history of every individual atom failure to do so represents not an inherent indeterminism in nature but rather a lack of obtainable information—and therefore a lack of understanding—concerning the mechanism of radioactive decay To mention but one possible alternative nonrelativistic quantum theory can be reinterpreted in terms of hidden variables which in principle determine the precise behavior of an individual system but whose values are not ascertained in measurements of the type which now can be carried out This alternative has not led to new predictions however and contains some unappealingly *ad hoc* features See CAUSALITY PROBABILITY IN PHYSICS

Wave function When the system is known to be in the eigenstate corresponding to $A = \alpha$ the eigenfunction $u(x,\alpha)$ is the wave function that is *it is the function whose projection on an eigenfunction $v(x\,\beta)$ of any observable B gives the probability of measuring $B = \beta$* The wave function $\psi(x)$ may be known exactly in other words the state of the system may be known as exactly as possible (within the limitations of uncertainty and complementarity) even though $\psi(x)$ is not the eigenfunction of a known operator This circumstance arises because the wave function obeys Schrodinger's wave equation Knowing the value of $\psi(x)$ at time $t = 0$ the wave equation completely determines $\psi(x)$ at all future times In general however if $\psi(x\,0) = u(x\,\alpha)$ that is if $\psi(x\,t)$ is an eigenfunction of A at $t = 0$ then $\psi(x\,t)$ will not be an eigenfunction of A at later times $t > 0$ For example suppose at $t = 0$ a free particle (a particle moving under no forces) is

of the uncertainties in position and momentum is as small as possible $(\Delta x)_0(\Delta p_x)_0 \cong \hbar$ compare Eq (1a) Then it can be proved that $(\Delta x)_t$ the uncertainty in x at time t satisfies

$$(\Delta x)_t = \left[(\Delta x)_0^2 + \frac{t^2}{m^2} (\Delta p_x)_0^2 \right]^{1/2} \qquad (8)$$

where m is the particle mass Equation (8) is readily interpreted The root mean square spread at time t results from $(\Delta x)_0$ and from an uncertainty in the distance the particle has traveled the latter uncertainty is $t(\Delta v_x)_0 = t(\Delta p_x)_0/m$ If $(\Delta x)_0 = 0$ meaning $\psi(x\,0)$ is an eigenfunction of the x operator $(\Delta x)_t$ is infinite showing that $\psi(x\,t)$ cannot be an eigenfunction of the x operator When the particle is free the projections of the wave function on the eigenfunctions of the mo-

other atoms and molecules, as well as with neutron beams. In all cases, the observations agree with quantum mechanical expectation. Refinements of the experiment yield accurate measurements of the parameters characterizing the magnetic interactions of atoms, molecules, and atomic nuclei. See MAGNETIC RESONANCE; MOLECULAR BEAMS; NUCLEAR MOMENTS.

Atomic spectra. Spectroscopy, especially the study of atomic spectra, probably provides the most detailed quantitative confirmation of quantization. Granting that the energy in a light wave is quantized, it follows from conservation of energy and Eq. (2b) that when an atom emits a photon of frequency f, its initial energy E_i and final energy E_f after emission are related by the equation

$$E_i - E_f = hf \qquad (3a)$$

Similarly, when a photon of frequency f is absorbed, the initial energy E'_i and final energy E'_f after absorption, satisfy

$$E'_f - E'_i = hf \qquad (3b)$$

In quantum mechanics, as in classical mechanics, an electron remains in the vicinity of a proton, that is, it is "bound" to the proton, when and only when the magnitude of its kinetic energy T is less than the magnitude of its negative potential energy V (V is set equal to zero at infinite separation). Thus, the total energy of a stable hydrogen atom necessarily is negative; in other words, the only allowed energies of atomic hydrogen are $-R/n^2$, where n is an integer and $R = -13.6$ ev. Consequently the radiation emitted by atomic hydrogen must consist of a discrete set of frequencies, or lines, obeying the relation from Eq. (3a),

$$f = \frac{R}{h}\left(\frac{1}{n^2} - \frac{1}{m^2}\right) \qquad (4)$$

where m and n are integers, $m > n$. This simple argument provides a convincing explanation of the observation that atomic hydrogen has a line spectrum and not a continuous spectrum; in other words, it radiates discrete frequencies rather than a continuous band of frequencies. The observed lines very accurately satisfy Eq. (4); moreover, except for small relativistic and quantum field theory effects, the nonrelativistic Schrödinger equation accurately predicts not merely the frequencies of the lines, but also their relative intensities and widths. Because, on the average, the levels E_i and E_f of Eq. (3a) endure only for a limited time, their energies are uncertain by an amount ΔE, as explained in the earlier discussion of the uncertainty principle; therefore, the observed lines have a width not less than $\Delta f \cong h^{-1}(\Delta E_i + \Delta E_f)$ according to Eq. (3a). Heavier atoms have more complicated line spectra than hydrogen, and the frequencies they emit cannot be described by formulas as simple as Eq. (4). For these atoms, the agreement between experiment and the predictions of the nonrelativistic Schrödinger equation, though always very good, is not as precise as for hydrogen. All the evidence indicates, however, that the discrepancies arise because the Schrödinger equation cannot be solved exactly in many-electron atoms, so that the theoretical predictions necessarily are approximate. If approximations were not necessary (for example, if perfect computing machines were available), there is every reason to think that the predictions of the Schrödinger equation would be exactly correct, except for the aforementioned small relativistic and field theory effects.

Probability considerations. The uncertainty and complementarity principles, which limit the experimenter's ability to describe a physical system, must limit equally the experimenter's ability to predict the results of measurement on that system. Suppose, for instance, that a very careful measurement determines that the x coordinate of a particle is precisely $x = x_0$. This is permissible in nonrelativistic quantum mechanics. Then formally the particle is known to be in the eigenstate corresponding to the eigenvalue $x = x_0$ of the x operator. Under these circumstances, an immediate repetition of the position measurement again will indicate that the particle lies at $x = x_0$; if the particle is moving in a one-dimensional force field, described by the potential $V(x)$, the particle's potential energy will be exactly $V(x_0)$. Knowing that the particle lies at $x = x_0$ makes the momentum p_x of the particle completely uncertain, however, according to Eq. (1a). A measurement of p_x immediately after the particle is located at $x = x_0$ could yield any value of p_x from $-\infty$ to $+\infty$; a measurement of $T = p_x^2/2m$ could yield any value from 0 to $+\infty$, and in fact, the average or expectation value of T in these circumstances would be infinite.

More generally, suppose the system is known to be in the eigenstate corresponding to the eigenvalue α of the observable A. Then for any observable B which is to some extent complementary to A, that is, for which an uncertainty relation of the form of Eqs. (1) limits the accuracy with which A and B can simultaneously be measured, it is not possible to predict which of the many possible values $B = \beta$ will be observed. However, it is possible to predict the relative probabilities $P_\alpha(\beta)$ of immediately thereafter finding the observable B equal to β, that is, of finding the system in the eigenstate corresponding to the eigenvalue $B = \beta$. If the system is prepared in the eigenstate α of A a great many times, and each time the observable B is measured immediately thereafter, the average of these observed values of B will equal the expectation value of B, defined as

$$\langle B \rangle = \Sigma \beta P_\alpha(\beta) \qquad (5)$$

summed over all eigenvalues of B; when the spectrum is continuous, the summation sign is replaced by an integral. To the eigenvalues correspond eigenfunctions, in terms of which $P_\alpha(\beta)$ can be computed. In particular, when α is a discrete eigenvalue of A and the operators depend only on x and p_x, the probability $P_\alpha(\beta)$ is postulated to be

energies F of Eq (15) are the energies of the stationary states (states not changing with time) of the system (*see* STATIONARY STATE)

The forms of Eqs (11*a*), (13), and (15) suggest that the classical observable p_x must be replaced by the operator $(\hbar/i)(\partial/\partial x)$ With this replacement

$$(xp_x - p_x x)\psi = i\hbar\psi \qquad (16)$$

In other words whereas the classical canonically conjugate variables x and p_x are numbers obeying the commutative law

$$xp_x - p_x x = 0 \qquad (17a)$$

the quantum mechanical quantities x and p_x are noncommuting operators obeying

$$xp_x - p_x x = i\hbar \qquad (17b)$$

In the formal theory the noncommutativity of x p_x leads directly to the uncertainty relation (1*a*) For a derivation of Eq (1*a*) from Eq (17*b*) as well as for generalizations and more sophisticated derivations of Eqs (11) (13) and (17*b*), *see* MATRIX MECHANICS QUANTUM THEORY NONRELATIVISTIC *See also* HAMILTON S EQUATIONS OF MOTION

Correspondence principle Since classical mechanics and Maxwell s electromagnetic theory accurately describe macroscopic phenomena quantum mechan . . .

it is

thou

for . . . complicated quantum mechanical systems its validity is well established by numerous illustrations such as those mentioned in the preceding discussions of the uncertainty principle and wave particle duality In general the classical limit is approached when (i) $\hbar \to 0$ (ii) the mass becomes large (iii) wavelengths become small (iv) dimensions become large (v) quantum numbers become large (The notation $\hbar \to 0$ refers to a mathematical operation in which one adheres to a fixed set of values of the other quantities involved and considers the effect of making $\hbar$ smaller and smaller) These simple criteria must be employed cautiously since they have not been stated in terms of dimensionless parameters obviously the classical limit in (ii) for instance cannot depend on the unit of mass Nonetheless these criteria are useful guides for example Eq (17*a*) is the limit of Eq (17*b*) as $\hbar \to 0$

Before the introduction of the Schrodinger equation (1926) made possible exact determination of energy levels and related quantum numbers the correspondence principle was very effectively employed as a heuristic means of arriving at the quantization rules In particular for periodic orbits there was evolved the rule

$$\oint p\,dq = nh \qquad (18)$$

where n is an integer q and p are canonically conjugate position and momentum variables, and the integration is performed along the orbit for one complete cycle *See* HAMILTON JACOBI THEORY, PERTURBATION (QUANTUM MECHANICS)

Equation (18) is not always exact but nonetheless is useful especially in the limit of high quantum numbers In the case of a one dimensional harmonic oscillator for example whose classical frequency is f and whose potential energy is $V(x) = 2\pi^2 mf^2x^2$ Eq (18) implies that the allowed energies are

$$E = nhf \qquad (19a)$$

whereas the correct result deduced from the Schrödinger equation is

$$E = (n + \tfrac{1}{2})\ hf \qquad (19b)$$

(*see* HARMONIC OSCILLATOR) The existence of the zero-point energy $\tfrac{1}{2}\ hf$ in the ground state $n = 0$ has been confirmed in analyses of molecular spectra and has significance for many phenomena, for example *see* IONIC CRYSTALS

Quantum statistics, indistinguishability Equation (18) means that the allowed orbits plotted as functions of q and p have area nh, the area between two allowed orbits equals h Since in classical statistical mechanics (i) all orbits are allowed, and (ii) the statistical weight of a volume $dq\,dp$ of q p (phase) space is proportional to $dq\,dp$ the correspondence principle suggests that in quantum statistics each allowed orbit replaces a set of classical orbits which cover an area h when plotted in the q p plane (see BOLTZMANN STATISTICS, QUANTUM STATISTICS STATISTICAL MECHANICS) It follows that in quantum statistics (i) each allowed quantized orbit should have the same statistical weight which may be set equal to unity (ii) with unit weight for a quantized orbit the weight of an unquantized volume $dp\,dq$ of phase space must be $(dp\,dq)/h$ Because the average number of atoms in a state of energy E is proportional to $\exp(-E/kT)$ where k is Boltzmann s constant and T the absolute temperature it is reasonable that quantum statistics yields different predictions than classical (Boltzmann) statistics when the energy level spacing is large compared to kT However in the limit that the level spacing becomes small compared to kT which corresponds to the limit $h \to 0$ quantum statistics must be equivalent to Boltzmann statistics The quantum theory of specific heat and the theory of black body radiation illustrate and confirm the considerations of this paragraph *See* HEAT RADIATION, SPECIFIC HEAT OF SOLIDS.

fac

Cl . . . in their ground states are placed in a box the statistic . . .

. . . rejects the possibility of this identification and simultaneously modifies the statistical weights which otherwise would be used in

mentum operator do not change with time corresponding to the classical result that the momentum of a free particle does not change Thus the probability of measuring any value of the momentum does not change $(\Delta p_x)_t = (\Delta p_x)_0$ and Eq (8) shows that $(\Delta x)_t (\Delta p_x)_t$ grows with time for a free particle whatever the value of $(\Delta x)_0$ provided $(\Delta x)_0 (\Delta p_x)_0 \cong \hbar$ Nonetheless $\psi(x\,t)$ is known no less exactly than the initial wave function $\psi(x\,0)$ The magnitude of $\Delta x\, \Delta p_x$ at any instant is no measure of the exactness with which the state of the system is known increased uncertainties *in position or momentum may be the price* for increased certainty in the value of some other observable

A system described by a wave function is said to be in a pure state Not all systems are described by wave functions however Consider for example a beam of hydrogen atoms streaming in the x direction out of a small hole in a hydrogen discharge tube According to the formal theory if the beam were described by a wave function $\psi(x)$ then

$$\psi(x) = C\ (x)\ u\ (z) + C\ (x)\ u\ (z) \qquad (9)$$

where $u^+(z)$ is the eigenfunction corresponding to finding a hydrogen atom with its z component J_z of total angular momentum equal to $\hbar/2$ $u\ (z)$ is the corresponding eigenfunction for finding $J_z = -\hbar/2$ $|C\ (x)|$ is the probability of finding $J_z = \hbar/2$ at any point x along the beam $|C\ (x)|^2$ is the corresponding probability of finding $J_z = -\hbar/2$ Since there are only two possibilities $J_z = \pm\hbar/2$ $|C\ (x)| + |C\ (x)|^2 = 1$ and since there seems no reason to favor either of these possibilities it is reasonable to suppose that $|C\ (x)| = |C\ (x)|^2 = ½$ As Eq (9) shows however to specify $\psi(x)$ it *is necessary to know not merely the relative mag*nitudes of the complex numbers $C\ (x)$ and $C\ (x)$ but also their relative phase It can be shown that each choice of relative magnitude and phase of C and C corresponds to a direction γ for which there is probability 1 of finding $J\ -\hbar/2$ and *probability zero of finding* $J_\gamma = -\hbar/2$ (*see* TRANSFORMATION THEORY QUANTUM) Thus each choice of relative magnitude and phase of C and C puts the system in an eigenfunction $u\ (\gamma)$ or $u\ (\gamma)$ that is in a pure state On the other hand the discharge tube singles out no particular direction in space so that in a Stern Gerlach experiment the original beam must break up into two beams of equal intensity whatever the direction of the external magnetic field Consequently the original beam is not in a pure state but can be regarded as a statistical ensemble or mixture of pure states oriented with equal probability in all directions Equivalently Eq (9) can be used for the original beam provided calculations are averaged over all relative phases of C and C The distinction between mixtures and pure states is strongly analogous to the distinction between polarized and unpolarized light beams consequently beams of particles in pure spin states are termed polarized

Schrodinger equation The expression

$$\psi(x\,t) = A(\lambda)\ exp\left[2\pi\left(\frac{x}{\lambda} - ft\right)\right] \qquad (10)$$

describes a plane wave of frequency f wavelength λ and amplitude $A(\lambda)$ propagating in the positive x direction (*see* WAVE MOTION) The previous discussion concerning wave particle duality suggests that this is the form of the wave function for a beam of free particles moving in the x direction with momentum $p = p_x$ with Eqs (2) specifying the connections between f λ and E p Differentiating Eq (10) it is seen that

$$p_x\psi = \frac{h}{\lambda}\psi - \frac{\hbar}{i}\frac{\partial\psi}{\partial x} \qquad (11a)$$

$$E\psi = hf\psi = -\frac{\hbar}{i}\frac{\partial\psi}{\partial t} \qquad (11b)$$

Since for a free particle $E = p\ /2m$ it follows also that

$$\frac{-\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} = -\frac{\hbar}{i}\frac{\partial\psi}{\partial t} \qquad (12)$$

Equation (12) holds for a plane wave of arbitrary λ and therefore for any superposition of waves of arbitrary λ that is arbitrary p_x Consequently Eq (12) should be the wave equation obeyed by the wave function of any particle moving under no forces whatever the projections of the wave function on the eigenfunctions of p_x Equations (11) and (12) further suggest that for a particle whose potential energy $V\ (x)$ changes *in other* words for a particle in a conservative force field $\psi(x\,t)$ obeys

$$\frac{-\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi = -\frac{\hbar}{i}\frac{\partial\psi}{\partial t} \qquad (13)$$

(*see* FORCE) Equation (13) is the time dependent Schrodinger equation for a one dimensional (along x) spinless particle Noting Eq (11*b*) and observing that Eq (13) has a solution of the form

$$\psi(x\,t) = \psi(x)\ \ \exp\ (-iEt/\hbar) \qquad (14)$$

it is inferred that $\psi(x)$ of Eq (14) obeys the time independent Schrodinger equation

$$\frac{-\hbar^2}{2m}\frac{\partial^2\psi}{\partial x^2} + V(x)\psi = E\psi \qquad (15)$$

Equation (15) is solved subject to reasonable boundary conditions for example, that ψ must be continuous and must not become infinite as x approaches $\pm\infty$ These boundary conditions restrict the values of E for which there exist acceptable solutions $\psi(x)$ to Eq (15) the allowed values of E depending on $V\ (x)$ In this manner the allowed energies of atomic hydrogen listed in the earlier discussion of quantization are obtained When a $\psi(x)$ solving Eq (15) exists all probabilities inferred from the corresponding $\psi(x\,t)$ are independent of time see Eq (6) Thus the allowed

acter of the wave functions For antisymmetric wave functions only one particle (without spin) may occupy a state for symmetric wave functions any number is possible Based on this distinction there are two separate distributions the Fermi D rac distribution for systems which are described by antisymmetric wave functions and the Bose Einstein distribution for systems described by sym metric wave functions

In relativistic quantum theory it is shown that particles having integer spin necessarily obey Bose Einstein statistics while those having half integer spin necessarily obey Fermi Dirac statis-

n /D

strictly speaking no system is correctly described by this distribution In practice of course the Boltzmann distribution gives an exceedingly good

tum description is essential *See* BOSE EINSTEIN STATISTICS FERMI DIRAC STATISTICS *see also* BOLTZMANN STATISTICS EXCLUSION PRINCIPLE KINETIC THEORY OF MATTER QUANTUM MECHANICS QUANTUM THEORY NONRELATIVISTIC QUANTUM THEORY RELATIVISTIC SPIN (QUANTUM MECHANICS) STATISTICAL MECHANICS [M DR]

Quantum theory, nonrelativistic

The modern theory of matter and its interaction with electromagnetic radiation applicable to systems of material particles which move slowly compared to the velocity of light and which are neither created nor destroyed

This article details the formal structure of quantum theory which can be summarized in the form of postulates as unequivocal as are Newton s laws of classical mechanics a less logically rigorous presentation is adopted here however Even so the reader unfamiliar with quantum theory is advised to read first another article (*see* QUANTUM MECHANICS) which more qualitatively discusses the novel (from the standpoint of classical physics) features of nonrelativistic quantum theory

These two articles attempt to make convincing the thesis that the formalism of nonrelativistic quantum theory is an *unarbitrary* and physically reasonable extension of the older classical theories Belief in quantum theory stems as much from acceptance of this thesis as from the *broad range* barely hinted at in these articles of successful application of the theory For added details concerning special formal topics *see* MATRIX MECHANICS PERTURBATION (QUANTUM MECHANICS) SPINOR TRANSFORMATION THEORY QUANTUM For generalizations of nonrelativistic quantum theory to relativistic particles (particles with speeds close to the velocity of light) or to systems in which particle creation and destruction can occur *see* QUANTUM ELECTRODYNAMICS QUANTUM FIELD THEORY, QUANTUM THEORY RELATIVISTIC

WAVE FUNCTION AND PROBABILITY DENSITY

Basic to quantum mechanics is the belief that the wave properties of matter are associated with a function the wave function obeying an equation called a wave equation The simplest possible wave in three dimensional space is a so called scalar wave in which the wave disturbance is wholly characterized by a single function $\psi(x\,y\,z,t) \equiv \psi(\mathbf{r}\,t)$ It is natural therefore to postulate that a wave function $\psi(x,y,z,t)$ provides a complete description of the simplest possible physical system namely a single particle *moving in a force field* specified by a potential $V(\mathbf{r})$ It is further postulated that $|\psi(\mathbf{r},t)|$ which classically would be proportional to the wave *intensity is the probability* density that is $|\psi(\mathbf{r}\,t)|^2\,d\mathbf{r}$ is the probability at time t of finding the particle in the volume $dxdydz \equiv d\mathbf{r}$ of space lying between x and $x + dx$ y and $y + dy$ z and $z + dz$

There is the obvious generalization that a wave function $\psi(\mathbf{r}_1 \quad \mathbf{r}_g\,t)$ will completely describe a system of g particles with $|\psi(\mathbf{r}_1 \quad \mathbf{r}_g\,t)|^2\,d\mathbf{r}_1 \quad d\mathbf{r}_g$ the probability at time t of simultaneously finding particle 1 in the volume element $d\mathbf{r}_1 = dx_1\,dy_1\,dz_1$, particle g in $d\mathbf{r}_g$ Moreover for a system of g distinguishable particles the probability $P_j(\mathbf{r})d\mathbf{r}$ of finding particle j in the volume element $d\mathbf{r}$ at $\mathbf{r} \equiv x\ y,\ z$ of physical space is

$$P_j(\mathbf{r})\,d\mathbf{r} \quad d\mathbf{r}\int(d\mathbf{r}_1 \quad d\mathbf{r}_{j-1}\,d\mathbf{r}_{j+1} \quad d\mathbf{r}_g \times |\psi(\mathbf{r}_1 \quad \mathbf{r}_{j-1}\,\mathbf{r}\,\mathbf{r}_{j+1} \quad ,\mathbf{r}_g)|^2\} \tag{1}$$

wherein $|\psi(\mathbf{r}_1 \quad \mathbf{r}_g)|^2$ is integrated over all positions of particles 1 to $j - 1$ and $j + 1$ to g with $\mathbf{r}_j$ put equal to $\mathbf{r}$

Normalization Because each of the particles 1 g must be somewhere in physical space Eq (1) demands that

$$\int d\mathbf{r}\,P_j(\mathbf{r}) = \int d\mathbf{r}_1 \quad d\mathbf{r}_g|\psi(\mathbf{r}_1 \quad \mathbf{r}_g)|^2 = 1 \tag{2a}$$

integrated over all positions of all g particles When Eq (2a) is satisfied ψ is said to be normalized If $\psi'(\mathbf{r}_1 \quad \mathbf{r}_g\,t)$ is a proposed wave function which satisfies

$$\int d\mathbf{r}_1 \quad d\mathbf{r}_g|\psi'|^2 = C \neq 1 \tag{2b}$$

the probabilities specified by ψ are found from Eq (1) using the normalized $\psi = C^{-1/2}\psi' \exp(i\eta)$, provided C is not infinite that is provided ψ' is *quadratically integrable* the phase factor $\exp(i\eta)$ η being real can be chosen arbitrarily Though the absolute probabilities $P_j(\mathbf{r})$ are *not defined* when $\psi'(\mathbf{r}_1 \quad \mathbf{r}_g)$ is not quadratically integrable $|\psi'|^2$ may remain a useful measure of the relative probability of finding particles 1 , g at $\mathbf{r}_1 \quad \mathbf{r}_g$

One need be concerned only with wave functions which can represent actual physical systems. It is postulated that an admissible (physically possible) wave function $\psi'(\mathbf{r}_1 \quad \mathbf{r}_g)$ is quadratically inte

computing statistical averages Formally this modification is accomplished by insisting that for a system of identical particles the wave function must be symmetric in the coordinates of all the particles including spin coordinates specifying spin orientations if the particles have integral spin namely 0 $\hbar$ $2\hbar$ and antisymmetric in the coordinates of all the particles including spin coordinates if the particles have half integral spin namely $\hbar/2$ $3\hbar/2$ $5\hbar/2$ (*see* BOSE EINSTEIN STATISTICS FERMI DIRAC STATISTICS) A symmetric wave function is unchanged under interchange of coordinates for example for two particles $\psi(x_1 x\ ) = \cos(x_1 - x_2)$ is symmetric an antisymmetric wave function changes sign under interchange of coordinates for example $\psi(x_1 x_2) = \sin(x_1 - x_2)$ is antisymmetric It is postulated that wave functions having other symmetry properties than these specified here never occur and therefore must not be included in any enumeration of available states even though without these symmetry restrictions they might be acceptable solutions of the Schrodinger equation (for the many particle system)

These symmetry restrictions have profound consequences for macroscopic as well as microscopic systems and are extremely well established for nuclei as well as atoms The reason for the connection between spin and statistics is beyond the scope of this article but appears to be a consequence of requirements imposed by the special theory of relativity *See* EXCLUSION PRINCIPLE [E C]

Bibliography D Bohm *Quantum Theory* 1951 M Born *Atomic Physics* 6th ed 1957 R B Lindsay and H Margenau *Foundations of Physics* 1957 W Pauli (ed) *Niels Bohr and the Development of Physics* 1955 F K Richtmyer E H Kennard and T Lauritsen *Introduction to Modern Physics* 5th ed 1955 H Semat *Introduction to Atomic and Nuclear Physics* 3d ed 1954

Quantum numbers

The integral or half integral numbers which characterize the discrete energy states of an elementary particle or system of particles such as an atom or molecule In quantum mechanics their possible values are obtained by solving the wave equation for the appropriate system Since angular momentum and the component of angular momentum in a given direction are frequently restricted to quantized values many kinds of quantum numbers give a measure of the corresponding angular momentum in quantum units of $h/2\pi$ or $\hbar$ where h is Planck s constant

For atomic systems the important quantum numbers are called the principal quantum number the orbital quantum number the spin quantum number the total angular momentum quantum number and the magnetic quantum number

The principal quantum number n appears in the expression for the energies of the hydrogen atom which are proportional to $-1/n^2$ (*see* ATOMIC STRUCTURE AND SPECTRA) Possible values of n are 1 2 3 etc In the original Bohr theory n appeared as the sum of the radial and azimuthal quantum numbers measuring the corresponding momenta and was formerly referred to as the total quantum number

The orbital (azimuthal) quantum number l gives the orbital angular momentum of a single electron in a particular energy state It can be shown that this angular momentum is $\sqrt{l(l+1)}\ \hbar$ (*see* QUANTUM MECHANICS) If the state has principal quantum number n possible values of l are 0 1 2 $(n-1)$ Closely related to l is the quantum number L of the total orbital angular momentum of two or more electrons which is also sometimes called the azimuthal quantum number Electrons are denoted s p d etc for values of $l = 0$ 1 2 etc and atomic states S P D etc for L 0 1 2 etc

The spin quantum number S specifies the resultant angular momentum of the electron spins The spin of each electron is $s = \frac{1}{2}$ quantum units and possible values of S represent algebraic sums of the individual s s The spin quantum number determines the multiplicity of an atomic term this multiplicity being $2S + 1$

The quantum number of the total angular momentum J represents in LS coupling the quantized vector resultant of L and S It can be defined for an atomic state whether or not LS coupling is appropriate For a single electron this quantum number is represented by a small j

The magnetic quantum number M measures the component of J projected in a fixed direction such as that of an applied electric or magnetic field The symbols M_L M_S m_l and m_s represent the quantized components of L S l s respectively Any magnetic quantum number gives accurately the component of the corresponding angular momentum in units of $\hbar$

Other quantum numbers are necessary for a description of the energy states of molecules and atomic nuclei and for the characterization of the several elementary particles found in high energy nuclear physics *See* ELEMENTARY PARTICLE MOLECULAR STRUCTURE AND SPECTRA NUCLEAR STRUCTURE [F A J]

Quantum statistics

The statistical description of particles or systems of particles whose behavior must be described by quantum mechanics rather than by classical mechanics As in classical that is Boltzmann statistics the interest centers on the construction of appropriate distribution functions However whereas these distribution functions in classical statistical mechanics describe the number of particles in given (in fact finite) momentum and positional ranges in quantum statistics the distribution functions give the number of particles in a group of discrete energy levels In an individual energy level there may be according to quantum mechanics either a single particle or any number of particles This is determined by the symmetry ch

, p_g commute with each other any component of r_1 or p_1 commutes with all components of r_2 , r_g and of p p_g the x coordinate of any particle commutes with p_y and p_z of that particle and so on

Hermitian operators An operator A relevant to a given system of g particles is termed Hermitian if

$$\sum_j \int dr_1 \quad dr_g\, \xi_j^*(A\psi)_j = \sum_j \int dr_1 \quad dr_g\,(A\xi)_j^*\psi_j \qquad (8)$$

for all pairs of sufficiently well behaved quadratically integrable wave functions

$$\xi(r_1 \quad r_g) \qquad \psi(r_1 \quad r_g)$$

In Eq (8) the asterisk denotes the complex conjugate and the sum is over all components j of ξ $A\psi$ $A\xi$ ψ Evidently every particle coordinate x_1 y_1 z_1 x_g y_g z_g is a Hermitian operator as is any reasonably well behaved $f(r_1 \quad r_g)$ Recalling that quadratically integrable functions vanish at infinity integration by parts shows that every component of p_1 p_g is Hermitian as is any $f(p_1 \quad p_g)$ expressible as a power series in components of p_1 p_g for example in the simple one dimensional spinless case

$$\int_{-\infty}^{\infty} dx\, \xi^* \frac{\hbar}{i}\frac{\partial\psi}{\partial x} = -\frac{\hbar}{i}\int_{-\infty}^{\infty} dx\, \frac{\partial\xi^*}{\partial x}\psi = \int_{-\infty}^{\infty} dx \left(\frac{\hbar}{i}\frac{\partial\xi}{\partial x}\right)^*\psi \qquad (9)$$

It is implied that ξ and ψ are continuous, otherwise the integration by parts yields extra terms on the right side of Eq (9) Similarly p_x^2 has the desired Hermitian property (8) only when ξ and ψ are continuous and have continuous first derivatives at all points

When A B C are individually Hermitian

$$\int \xi^*[(ABC \quad)\psi] = \int [(\quad CBA)\xi]^*\psi \qquad (10)$$

For simplicity the integration variables and the summation over components are not indicated explicitly in Eq (10) When A and B are Hermitian Eq (10) implies (i) $\frac{1}{2}(AB + BA)$ is Hermitian (ii) AB and BA are not Hermitian unless A and B commute For example xp and $p\,x$ are not Hermitian but $\frac{1}{2}(xp_x + p_x x)$ is classically of course there is no distinction between xp_x $p\,x$ or $\frac{1}{2}(xp + p\,x)$ By appropriately symmetrizing taking n te of Eq (10) one can construct the quantum mechanical Hermitian operator corresponding to any classical $f(r_1 \quad r_g\ p_1 \quad p_g)$ expressible as a power series in components of coordinates and momenta

If one supposes that the forces acting on a quantum mechanical system of g particles are precisely the same as in the classical case the energy oper cal Hamiltonian

$$r_g\ p_1 \quad p_g) = T + V$$

where the kinetic energy $T = p_1^2/2m_1 + \quad + p_g^2/2m_g$, $V(r_1 \quad r_g)$ is the potential energy m is the mass of particle i (see HAMILTON'S EQUATIONS OF MOTION) Quantum mechanical nonclassical forces with associated potential energy operators more complicated than $V(r_1 \quad r_g)$ are not uncommon however For example the interaction between two neutrons is believed to include space exchange

$$V\psi_j(r_1 r_2) = J(r_{12})P_{12}\psi_j(r_1 r_2) = J(r_{12})\psi(r_2 r_1) \qquad (11)$$

where $J(r_{12})$ is an ordinary function of the distance r_{12} between the particles and the space exchange operator P_{12} interchanges the space coordinates of particles 1 and 2 in each component of ψ See NUCLEAR STRUCTURE

Real eigenvalues The eigenvalue equation for an operator A is

$$Au(\alpha) = \alpha u(\alpha) \qquad (12)$$

where the number α is the eigenvalue and the corresponding eigenfunction $u(\alpha)$ is a not identically zero wave function solving Eq (12) for that value of α The eigenvalue equation $Hu = Eu$ for the energy operator H has special importance and is known as the time-independent Schrodinger equation Since the eigenvalues α are identified with the [illegible]

form a complete set the meaning and importance of which is explained subsequently Property (i) is important because actual measurements yield real numbers only for example lengths, meter readings and the like If the eigenvalue of A were complex it could not be maintained that each value of α represented a possible result of exact measurement of A This assertion is not negated by the fact that it is formally possible to combine real quantities into complex expressions for example the coordinates of r in the xy plane form the complex vector $x + iy$

The eigenvalues α_n belonging to the quadratically integrable eigenfunctions $u(\alpha_n) \equiv u_n$ of a Hermitian operator A are necessarily real In the notation of Eq (10) letting $\xi = \psi = u_n$ in Eq (8) and employing Eq (12)

$$\alpha_n \int u_n^* u_n = \int u_n^*(\alpha_n u_n) = \int u_n^*(Au_n) = \int (Au_n)^* u_n = \int (\alpha_n u_n)^* u_n = \alpha_n^* \int u_n^* u_n \qquad (13)$$

The equality of the first and last terms of Eq (13) demonstrates that $\alpha_n = \alpha_n^*$ For this reason it is postulated that all observable" operators are Hermitian operators and conversely that all Hermitian operators represent observable quantities, henceforth all operators are supposed Hermitian In addition it is necessary to require that the allowed eigenfunctions preserve the hermiticity property (8) otherwise Eq (13) would not hold For the important class of Hamiltonian operators $H = T + V$ except for highly singular (discontinuous) potentials the boundary conditions that α and

grable (type 1), or fails to be quadratically integrable (type 2) only because ψ' vanishes too slowly or at worst remains finite as infinity is approached in the 3g dimensional space of $\mathbf{r}_1, \ldots, \mathbf{r}_g$. Convincing physical and mathematical reasons can be found for excluding wave functions other than these types. One particle wave functions $\psi'(\mathbf{r})$ of type 2 correspond to classically unconfined systems for example, an electron ionized from a hydrogen atom, for these systems $\psi(\mathbf{r}) = \infty^{-1/2}\psi' = 0$ has the obvious interpretation that an unconfined particle is sure to be found outside any given finite volume. Such a $\psi'(\mathbf{r})$ also can represent a very large (effectively infinite) number of independently moving identical particles in a very large (effectively infinite) volume, for example, a beam of free electrons issuing from an electron gun, in this event $|\psi'(\mathbf{r})|^2$, although it continues to describe the likelihood of observing an electron can be thought to equal the actual number density of electrons at any point $\mathbf{r}$, with $C = \infty$ in Eq. (2b) indicating that the number of particles in all of space is infinite. These considerations can be extended to quadratically nonintegrable many particle wave functions $\psi'(\mathbf{r}_1, \ldots, \mathbf{r}_g)$.

Spin The preceding formalism accepts the presumption of classical physics that a particle is a structureless entity about which "everything" is known when its position $\mathbf{r}$ is known. This presumption is inaccurate (and therefore the formalism is not wholly adequate) for systems of electrons protons and neutrons these being the fundamental particles which compose what usually is termed stable matter. In particular an electron, proton, or neutron cannot be described completely by a single wave function $\psi(\mathbf{r}\,t)$. For each of these particles two wave functions $\psi_1(\mathbf{r}\,t)$ and $\psi_2(\mathbf{r}\,t)$ are required which may be regarded as components of an over all two component wave function $\psi(\mathbf{r}\,t)$. The need for a multicomponent wave function has the immediate classical interpretation that the particle has internal degrees of freedom, that is that knowledge of the position of the particle is not "everything." It can be shown that these internal degrees of freedom are associated with so called intrinsic angular momentum or spin, *see* SPIN (QUANTUM MECHANICS). For electrons protons, or neutrons the spin is $\frac{1}{2}\hbar$, where $\hbar = h/2\pi$ and h is Planck's constant, the only allowed values of s_z, the z component of the spin, are $\pm\frac{1}{2}$ (in units of $\hbar$). Thus when the system contains a single parti

Eq. (2a) is

$$\int d\mathbf{r}\,[|\psi_1(\mathbf{r},t)|^2 + |\psi_2(\mathbf{r},t)|^2] = 1 \qquad (3)$$

This formalism is readily extended to many particle systems. For example, when the system contains g particles of spin $\frac{1}{2}$ the over all wave function ψ has 2^g components ψ_j, $|\psi_1(\mathbf{r}_1, \ldots, \mathbf{r}_g)|^2$ is the probability density for finding each of particles 1 to g with spin oriented along $+z$ and the normalization condition is

$$\sum_j \int d\mathbf{r}_1 \cdots d\mathbf{r}_g\,|\psi_j(\mathbf{r}_1, \ldots, \mathbf{r}_g)|^2 = 1 \quad (4)$$

summed from $j = 1$ to 2^g. The appropriate reinterpretations when the wave function is not quadratically integrable are obvious. Complications arising from particle indistinguishability are discussed later.

OPERATORS

Whereas in classical mechanics particle coordinates $\mathbf{r}$ and momenta $\mathbf{p}$ are numbers which can be specified independently at any instant of time in quantum mechanics the components of $\mathbf{r}$ and $\mathbf{p}$ are linear operators, as also are functions $f(\mathbf{r}\,\mathbf{p})$ of the coordinates and momenta. It is postulated that (i) the operator $\mathbf{x}$ (here distinguished by boldface from the x coordinate to which $\mathbf{x}$ corresponds) simply multiplies a wave function $\psi(x)$ by x that is $\mathbf{x}\psi = x\psi$, (ii) the operator corresponding to the canonically conjugate x component of momentum of that particle is $p_x = (\hbar/i)\,\partial/\partial x$, that is $p_x\psi = (\hbar/i)\,\partial\psi/\partial x$. Thus $A\psi$ denotes the new wave function $\psi' = A\psi$ resulting from the linear operation A on a given wave function ψ. When A, B, C are linear operators, and ψ, ξ are any two functions,

$$\begin{aligned} A(\psi + \xi) &= A\psi + A\xi \\ (A + B)\psi &= A\psi + B\psi \\ AB\psi &= A(B\psi) \end{aligned} \qquad (5)$$

Moreover, if ξ and ψ can be added or equated then ξ and ψ must have the same number of components and depend upon the same space and spin coordinates, if $\xi = \psi$ corresponding components of ξ and ψ are equal. A spin independent operator performs the same operation on each component of a many component wave function for example $(p_x\psi)_j = p_x\psi_j = (\hbar/i)\,\partial\psi_j/\partial x$. Spin dependent operators are more complicated, for instance in a one particle system of spin $\frac{1}{2}$ the components of $\psi' = s_z\psi$, where s_z denotes the z component of the spin operator are

$$\psi'_1 = \tfrac{1}{2}\hbar\psi_1 \qquad \psi'_2 = -\tfrac{1}{2}\hbar\psi_2$$

using the notation adopted previously in the discussion of spin.

The operators A and B are said to commute when for any ψ $A(B\psi) = B(A\psi)$ implying

$$AB - BA = 0 \qquad (6)$$

The operator $(AB - BA)$ is termed the commutator of A and B. Any operator $f(A)$ expressible as a power series in the operator A commutes with A. Performing the indicated operations

$$(xp_x - p_xx)\psi = i\hbar\psi \qquad (7)$$

Equation (7) shows that pairs of operators need not commute. In a g particle system all particle coordinates $x_1\ y_1,\ z_1 \ldots x_g\ y_g,\ z_g$ commute with each other, all momentum coordinates p

continuous spectrum, each $u(x\,k)$ can be regarded as representing a beam of particles all moving with the same velocity Because $\exp[ikx]$ is periodic with wavelength $\lambda = 2\pi/k$, such a beam will demonstrate wavelike properties, in fact $p_x = \hbar k = h/\lambda$ in agreement with the de Broglie relations This agreement is not trivial, although the conclusion that the quantum mechanical momentum operator is $p_x = (\hbar/i)\,\partial/\partial x$ can be argued starting from the de Broglie relations the form of p_x also can be inferred directly from Eq (7), which in turn can be argued from the formal analogy between the properties of the commutator and the Poisson bracket without any reference to the de Broglie relations *See* CANONICAL TRANSFORMATIONS.

Normalized on the k scale, the eigenfunctions are

$$u(x\,k) = \frac{1}{\sqrt{2\pi}}e^{ikx} \qquad (18a)$$

for which corresponding to Eq (15),

$$\int_{-\infty}^{\infty} dx\, u^*(k)u(k') = \frac{1}{2\pi}\int_{-\infty}^{\infty} dx\, e^{i(k'-k)x} = \delta(k-k') \qquad (18b)$$

Eq (18b) *amounts* to a nonrigorous statement of the Fourier integral transform theorem that when any sufficiently regular

$$\psi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dk\, e^{ikx}c(k) \qquad (19a)$$

then $$c(k) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\, e^{-ikx}\psi(x) \qquad (19b)$$

(*see* FOURIER SERIES AND INTEGRALS) Equation (19b) can be derived by mathematically rigorous procedures The quantum theoretic derivation multiplies Eq (19a) by $(2\pi)^{-1/2}\exp(-ik'x)$ and integrates over all x There is obtained *after interchanging* the orders of x and k integration

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\, e^{-ik'x}\psi(x)$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} dk\, c(k)\int_{-\infty}^{\infty} dx\, e^{i(k-k')x}$$
$$= \int_{-\infty}^{\infty} dk\, c(k)\,\delta(k-k') = c(k') \qquad (20)$$

Kinetic energy For the kinetic *energy* $T_x = p_x^2/2m$ Equation (12) is

$$T_x u = -\frac{\hbar^2}{2m}\frac{\partial^2 u}{\partial x^2} = Eu \qquad (21)$$

The eigenvalue is E Admissible solutions are

$$u(x\,E) = C(E)\exp\left(\frac{i}{\hbar}x\sqrt{2mE}\right) \qquad (22)$$

where $\sqrt{E}$ must be real Thus the spectrum is continuous and runs from $E = 0$ to $E = +\infty$ The square root may be either positive or negative in Eq (22) so that there are two independent eigen ... of E that is the spectrum ... led

$$u_\pm(x,E) = \left(\frac{m}{2h^2E}\right)^{1/4}\exp\left(\pm\frac{i}{\hbar}x\sqrt{2mE}\right) \qquad (23)$$

In Eq (23) *the square* root always is positive, u_+ and u_- individually satisfy Eq (15), but the sets u_+ and u_- are orthogonal to each other Introducing $k = \sqrt{2mE}/\hbar$ the eigenfunctions can be labeled by the single *parameter* k, $-\infty \leqq k \leqq \infty$ instead of by E, $0 \leqq E \leqq \infty$, and the subscripts + or − Evidently the $u(x\,k)$ of Eq (18a) are eigenfunc*tions not only of* p_x *but also of* T_x, each $u(x,k)$ corresponds to a different eigenvalue of p_x, but $u(x,k)$ and $u(x\,-k)$ correspond to the same eigenvalue of T_x Interpreted physically, these results *mean that the energy* E *of a free particle is known* exactly when its momentum p_x is *known exactly*, and that $E = p_x^2/2m = (-p_x)^2/2m$ just as in classical mechanics The normalizations (18a) and (23) are different because $dE = (\hbar^2k/m)\,dk$ The eigenfunctions

$$u'_+(x\,E) = \frac{1}{\sqrt{2}}[u_+(x,E) + u_-(x\,E)]$$
$$= \left(\frac{2m}{h^2E}\right)^{1/4}\cos\frac{x}{\hbar}\sqrt{2mE} \qquad (24a)$$

$$u'_-(x\,E) = \frac{1}{\sqrt{2}}[u_+(x,E) - u_-(x,E)]$$
$$= \left(\frac{2m}{h^2E}\right)^{1/4}\sin\frac{x}{\hbar}\sqrt{2mE} \qquad (24b)$$

also are normalized on the E scale, and are an al*ternative set* to $u(x,E)$ of Eq (22)

Typical energy operator. Equation (12) is

$$Hu = (T_x + V)u = \left[\frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]u = Eu \qquad (25)$$

where the operators in the brackets operate on u Typically but not always (see the two subheadings immediately following) $V(x)$ is an everywhere smooth finite function which approaches zero at $x = \pm\infty$ Figure 1 is a plot of such a $V(x)$, having attractive force *centers* (negative potential energy) near $x = \pm a$ and a *repulsive* region (positive potential energy) near $x = 0$ the minimum potential at $x = \pm a$ is $V_0 < 0$

Rewrite Eq (25) in the form

$$\frac{\partial^2 u}{\partial x^2} = Ku \qquad (26a)$$

$$K(x\,E) = \frac{2m}{\hbar^2}[V(x) - E] \qquad (26b)$$

When $K > 0$ Eq (26b) implies that $\partial/\partial x(\partial u/\partial x)$ has the same sign as u, for example $\partial u/\partial x$ increases with increasing x if $u > 0$ In other words at points x where $K(x\,E) > 0$, *solutions* of Eq (26a) are convex toward the x axis, where $K(x,E) < 0$, $u(x,E)$ is concave toward the x axis Provided $V(x)$ *decreases* to zero sufficiently rap

$\partial u/\partial x$ must be continuous guarantee reality of the eigenvalue corresponding to a quadratically integrable u solving Eq (12) Admissible wave functions defined following Eq (2b), may be quadratically nonintegrable, however It always is assumed that the physically desirable properties (i) and (ii) that were discussed in the preceding paragraph follow from the equally physically desirable simple requirement that the eigenfunctions $u(\alpha)$ of A must be admissible provided that u, $\partial u/\partial x$, etc, satisfy the continuity conditions which make Eq (13) correct for quadratically integrable u_n In systems containing a single spinless particle this assumption has been justified rigorously for operators of interest, the widespread quantitative successes of quantum theory support the belief that the assumption is equally valid in more complicated systems

Orthogonality. The eigenvalues α corresponding to quadratically integrable eigenfunctions typically form a denumerable (countable) though possibly infinite set, and compose the discrete spectrum of A The admissible nonquadratically integrable eigenfunctions typically correspond to a continuous set of real eigenvalues composing the continuous spectrum of A An eigenvalue α is degenerate with order of degeneracy $d \geqq 2$, if there exist d independent eigenfunctions $u_1, \ldots, u_d$ corresponding to the same value of α, whereas every $d+1$ such eigenfunctions are dependent The d functions $\psi_1, \ldots, \psi_d$ are (linearly) dependent if the equation

$$c_1\psi_1 + \cdots + c_d\psi_d = 0 \qquad (14)$$

can be true with not all the constants $c_1, \ldots, c_d$ equal to zero Eigenvalues which are either discrete, or continuous or both, may be degenerate While so called accidental degeneracy can occur, degeneracy of the eigenvalues α of A ordinarily is associated with the existence of one or more operators which commute with A In the absence of degeneracy the eigenfunctions $u(\alpha)$ are uniquely indexed by α and can be chosen to satisfy the orthonormal (orthogonality and normalizing) relations [compare Eq (4)]

$$\sum_j \int d\mathbf{r}_1 \cdots d\mathbf{r}_g\, u_j^*(\alpha)u_j(\alpha') = \delta_{\alpha\alpha'} \text{ or } \delta(\alpha - \alpha') \qquad (15)$$

In Eq (15) the Kronecker symbol $\delta_{\alpha\alpha'}$ is employed when α lies in the discrete spectrum, $\delta_{\alpha\alpha'} = 0$ for $\alpha \neq \alpha'$, $\delta_{\alpha\alpha'} = 1$ for $\alpha = \alpha'$ The Dirac delta function $\delta(\alpha - \alpha')$ is employed when α lies in the continuous spectrum, $\delta(\alpha - \alpha') = 0$ when $\alpha \neq \alpha'$, but has a finite integral namely

$$\int_{-\infty}^{\infty} d\alpha\, \delta(\alpha - \alpha') = \int_{-\infty}^{\infty} d\alpha'\, \delta(\alpha - \alpha') = 1 \qquad (16a)$$

For a wide variety of functions $f(\alpha)$ these properties of $\delta(\alpha - \alpha')$ imply that

$$\int_{-\infty}^{\infty} d\alpha\, f(\alpha)\, \delta(\alpha - \alpha') = \int_{-\infty}^{\infty} d\alpha\, f(\alpha)\, \delta(\alpha' - \alpha) = f(\alpha') \qquad (16b)$$

When Eq (15) holds, $u(\alpha)$ are said to be normalized on the α scale Evidently $\delta(x)$ is highly singular at $x = 0$ and is an even function of x

Equation (15) asserts that eigenfunctions corresponding to different eigenvalues always are orthogonal that is, that the integral in Eq (15) equals zero whenever $\alpha \neq \alpha'$ For discrete α (or α') this assertion is readily proved by an argument similar to Eq (13), in fact, the orthogonality for $\alpha \neq \alpha'$ holds whether or not the eigenvalues are degenerate When α and α' both lie in the continuous spectrum, however, the integral of Eq (15) does not converge and therefore actually is not defined. Thus the delta function $\delta(\alpha - \alpha')$ primarily is a useful formal device, here and elsewhere in the theory delta functions always are eventually integrated over their arguments as in Eqs (16), such integration makes expressions like the left side of Eq (15) effectively convergent A mathematically rigorous justification of the use of the delta function in quantum theory, or a rigorous justification of Eq (15) encounters difficulties These are related to the difficulties in establishing rigorously the aforementioned properties (i) and (ii) For operators and eigenfunctions of physical interest experience suggests no reason to doubt the basic correctness of the mathematical procedures of nonrelativistic quantum theory *See* OPERATOR THEORY

ILLUSTRATIVE APPLICATIONS

The immediately following subheadings apply the preceding formalism to some representative problems involving a one dimensional spinless particle free to move in the x direction only, in every case one begins by seeking the appropriate solution to Eq (12) As subsequent discussion makes clear, results for this simplest of systems are pertinent to more complicated systems

Momentum. Equation (12) is

$$p_x u = \frac{\hbar}{i}\frac{\partial u}{\partial x} = \hbar k u \qquad (17)$$

where $\hbar k$ is the eigenvalue, conventionally the eigenfunctions are indexed by the wave number k Solutions to Eq (17) have the form $u(x\,k) = C(k) \exp[ikx]$ where $C(k)$ is a normalizing constant When k is real $u(x\,k)$ is finite for all x $-\infty \leqq x \leqq \infty$ Consequently the spectrum is continuous and includes all real k, $-\infty \leqq k \leqq \infty$ If k has an imaginary part that is if $k = k_1 + ik_2$ $k_2 \neq 0$, then $u(x, k_1 + ik_2)$ is not admissible since it becomes infinite at either $x = +\infty$ or $x = -\infty$ If the eigenfunction could vanish identically for $|x| \geqq a$ it would be quadratically integrable for complex k, in this event the eigenfunction would be discontinuous at $x = a$, however since

$$|C(k) \exp[ikx]| \neq 0$$

unless $C(k) = 0$ Thus the requirements that u be admissible and continuous ensure that the eigenvalues of p_x are real as asserted previously in the discussion of real eigenvalues The $u(x\,k)$ are quadratically nonintegrable as expected for the

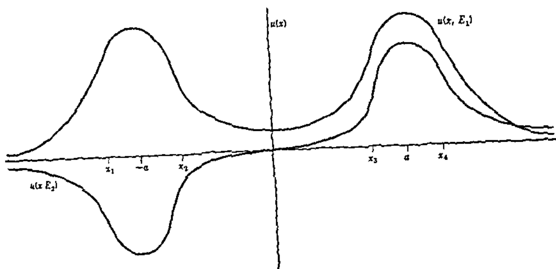


Fig 3 Eigenfunctions for the two lowest eigenvalues of the Hamiltonian with the potential of Fig 2

because it is insufficiently concave to join smoothly with L this makes understandable the general result that eigenvalues corresponding to quadratically integrable eigenfunctions are discrete

Potential barrier Consider Eq (25) with $V(x) = 0$ for $x < 0$ $V(x) = V_1 > 0$ for $x > 0$ and $0 < E < V_1$ as shown in Fig 4a As previously explained u and $\partial u/\partial x$ must be everywhere continuous even though $V(x)$ is discontinuous at $x = 0$ Recollecting Eqs (23) (26) and (27) for $0 < E < V_1$ there must be one and only one independent eigenfunction having the form

$$u(x\,E) = \exp\,[ix(2mE)^{1/2}/\hbar] + R\exp\,[-ix(2mE)^{1/2}/\hbar] \qquad x < 0 \quad (28a)$$

$$u(x\,E) = S\exp\,[-x(2m|E - V_1|)^{1/2}/\hbar] \qquad x > 0 \quad (28b)$$

In Eq (28) R and S are constants It is reasonable and customary to interpret the first exponential in Eq (28a) as a beam of particles moving with momentum $p_x = (2mE)^{1/2}$ toward the classical barrier or turning point $x = 0$, the second exponential in Eq (28a) represents a reflected beam Since the amplitude of the incident beam has been normalized to unity [not a normalization consistent with Eq (15)] $|R|^2$ must be the reflection coefficient of the barrier The continuity requirements at $x = 0$ yield

$$R = \frac{iE^{1/2} + (V_1 - E)^{1/2}}{iE^{1/2} - (V_1 - E)^{1/2}} \qquad S = \frac{2iE^{1/2}}{iE^{1/2} - (V_1 - E)^{1/2}} \qquad (29)$$

showing that $|R|^2 = 1$ in agreement with classical expectation for $E < V_1$ (see REFLECTION AND TRANSMISSION COEFFICIENTS) The eigenfunction $u(x,E)$ is sketched in Fig 4a Since $|u|^2 \neq 0$ at $x > 0$ particles penetrate the classically inaccessible region to the right of the barrier, because $|u|^2 = 0$ at $x = +\infty$, all particles eventually are turned back, however, consistent with $|R|^2 = 1$ This penetration, and Eqs (28) and (29), closely resembles the penetration (and corresponding classical optics expressions) of a totally reflected light wave into a medium with smaller index of refraction Hence Eqs (28) and (29), which are unforced results of solving Eq (25) for $V(x)$ of Fig 4a, manifest the wave particle duality inherent in the quantum theoretic formalism

For the barrier of finite thickness a (Fig 4b) there are two solutions at every $0 < E < V_1$ The solution representing a beam incident from the left, transmitted with coefficient $|T|^2$ and reflected with coefficient $|R|^2$, is found from Eq (28a) together with

$$u(x,E) = C\exp\,[-x(2m|E - V_1|)^{1/2}/\hbar] + D\exp\,[x(2m|E - V_1|)^{1/2}/\hbar] \qquad 0 < x < a \quad (30a)$$

$$u(x,E) = T\exp\,[ix(2mE)^{1/2}/\hbar] \qquad x > a \quad (30b)$$

Equations (30) mean that, except for the incident $\exp\,[ix(2mE)^{1/2}]$ at $x = -\infty$, the solution at $x = \pm\infty$ must consist of waves traveling *out* toward infinity, a similar outgoing boundary condition specifies the continuum solution $E > 0$ representing more complicated collisions for example, the scattering of a beam of particles by target particles in a foil (see SCATTERING EXPERIMENTS, ATOMIC AND MOLECULAR, SCATTERING EXPERIMENTS NUCLEAR) Outgoing boundary conditions are employed in classical wave theories as well See SCATTERING (ELECTROMAGNETIC RADIATION)

The continuity requirements lead to

$$|T|^2 = \left\{1 + \frac{V_1^2\sinh^2\,[(a/\hbar)\sqrt{2m(V_1 - E)}]}{4E(V_1 - E)}\right\}^{-1} \qquad (31a)$$

and $|R|^2 = 1 - |T|^2$, as is necessary if $|R|^2$ and $|T|^2$ are to represent respectively reflection and

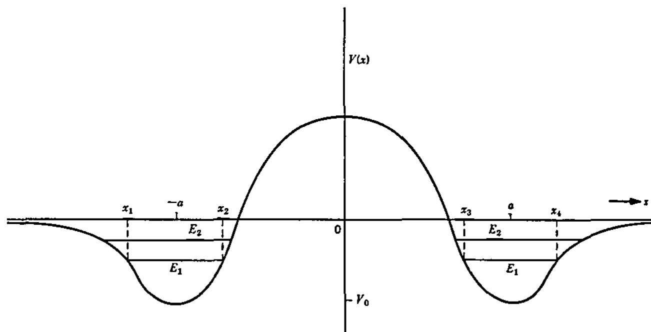


Fig 1 Illustrating a potential $V(x)$, capable of trapping particles near $x = \pm a$

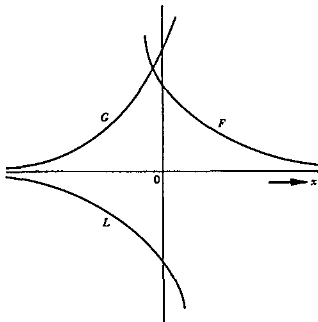


Fig 2 Everywhere convex solutions asymptotic to x axis

idly, $K(\pm\infty,E) < 0$ when $E > 0$, so that at $x = \pm\infty$ solutions to Eq (25) are approximated by linear combinations of the oscillatory functions of Eq (24), when $E < 0$, however, $K(\pm\infty,E) > 0$, so that at $x = \pm\infty$ solutions to Eq (25) are approximated by linear combinations of the convex functions [compare Eq (22)]

$$\begin{aligned} u_1(x,E) &= \exp\,[x(2m|E|)^{1/2}/\hbar] \\ u_2(x,E) &= \exp\,[-x(2m|E|)^{1/2}/\hbar] \end{aligned} \tag{27}$$

Function u_1 is infinite at $x = +\infty$, u_2 is infinite at $x = -\infty$ Thus, for $E < 0$, every admissible solution of Eq (25) must behave like u_1 at $x = -\infty$ and like u_2 at $x = +\infty$, that is, must be asymptotic to the x axis at both $x = +\infty$ and $x = -\infty$ When $V_0 < E < 0$ there are values of x where the solution is concave, so that a smooth (satisfying the boundary condition that u and $\partial u/\partial x$ must be continuous) transition from a curve like G or L (Fig 2) on the left to F on the right may be possible When $E < V_0$ solutions to Eq (25) are everywhere convex and like F G, L of Fig 2, never are asymptotic to the x axis at both $x = \pm\infty$

It is concluded that (i) the continuous spectrum includes all positive values of E $0 \leqq E \leqq \infty$, and at each such E there are two independent, nonquadratically integrable (oscillatory at $x = \pm\infty$) eigenfunctions as in the previously discussed example of kinetic energy, (ii) there are no eigenvalues $E < V_0$, (iii) there may, but need not be discrete eigenvalues $V_0 < E < 0$ corresponding to quadratically integrable eigenfunctions The lowest eigenfunction for $E = E_1$ has the least region of concavity and therefore joins F of Fig 2 to G without crossing the x axis as in Fig 3 The eigenfunction $u(x\,E_2)$ corresponding to the next higher eigenvalue $E = E_2 > E_1$ has one node (one zero), decreases less rapidly at $x = \pm\infty$ than $u(x,E_1)$ and links F to L the next higher eigenfunction corresponding to $E_3 > E_2$ has two nodes again links F to G and so on

The horizontal lines in the neighborhood of $x = \pm a$ in Fig 1 show the energy levels of the two lowest eigenvalues E_1 E_2 The points x_1 x_2 where $E = V_1$, and where the curvature of $u(x,E_1)$ changes sign (Fig 3) are the turning points (velocity equals zero) of a classical particle oscillating with total energy E_1 in the attractive potential well at $x = -a$, x_3 x_4 are similar turning points near $x = a$ For $E_1 < E < E_2$ a solution $u(x,E)$ starting out as does F at $x = +\infty$ crosses the x axis but becomes negatively infinite at $x = -\infty$,

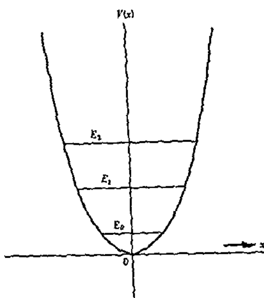
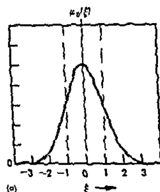
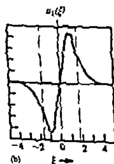


Fig 5 Harmonic oscillator potential





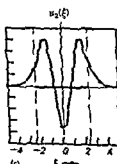


Fig 6 Harmonic oscillator eigenfunctions. (a) $n = 0$ (b) $n = 1$ (c) $n = 2$ (L Pauling and E B Wilson Jr Introduction to Quantum Mechanics McGraw Hill 1935)

genvalues of B are mutually orthogonal (see the earlier discussion on orthogonality) If all the β_i are not different a third operator C simultaneously commuting with A and B is sought and so on For the system of g particles of spin ½ one determines in this fashion a complete set of simultaneously commuting observables A B C whose corresponding eigenvalues α β γ (except for accidental degeneracy) uniquely index an orthonormal set of 2^g component eigenfunctions

$$u_j(\mathbf{r}_1, \quad \mathbf{r}_g, \alpha\,\beta\,\gamma, \quad), j = 1 \text{ to } 2^g$$

satisfying

$$\int u^*(\alpha\,\beta\,\gamma, \quad) u(\alpha',\beta',\gamma', \quad) = \delta(\alpha - \alpha')\ \delta(\beta - \beta')\ \delta(\gamma - \gamma') \qquad (35)$$

In Eq (35) and henceforth unless specifically in

bol $\delta_{\alpha\alpha'}$ when α α' are discrete and so on

The meaning of property (ii) mentioned in the earlier discussion of real eigenvalues can now be explained The eigenfunctions of an operator A having been indexed as in Eq (35) any sufficiently smooth quadratically integrable 2^g component wave function $\psi(\mathbf{r}_1 \quad x_g)$ can be expressed in the form

$$\psi(\mathbf{r}_1 \quad ,x_g) = \int d\alpha \int d\beta \int d\gamma \times c(\alpha\,\beta,\gamma \quad) u(\mathbf{r}_1, \quad \mathbf{r}_g, \alpha\,\beta\,\gamma, \quad) \qquad (36)$$

where each eigenvalue is integrated over its entire spectrum with the understanding (further to simplify the notation) that in the discrete spectrum integration is replaced by summation Employing Eq (35), the constants $c(\alpha,\beta,\gamma, \quad)$ are found to be

$$c(\alpha\,\beta,\gamma, \quad) = \int u_j^*(\alpha\,\beta\,\gamma, \quad)\psi \qquad (37)$$

Equations (36) and (37) are consistent with the often instructive interpretation that $c(\alpha\,\beta,\gamma \quad)$ are the projections of the vector ψ on a complete set of orthogonal unit vectors $u(\alpha\,\beta\,\gamma, \quad)$ in an infinite dimensional vector space

Probability considerations When ψ is normalized Eqs (4), (16), (35), and (36) imply

$$\int d\alpha \int d\beta \int d\gamma \qquad |c(\alpha,\beta,\gamma, \quad)|^2 = 1 \qquad (38)$$

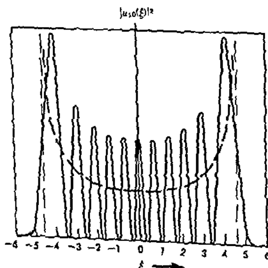


Fig 7 Probability density for harmonic oscillator in state $n = 10$ (L Pauling and E B Wilson, Jr Introduction to Quantum Mechanics McGraw Hill 1935)

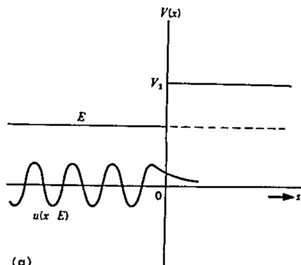
transmission coefficients Equation (31a) and the corresponding expression when $E > V_1$ resemble the equations for transmission of light through thin films When $(a/\hbar)\sqrt{2m(V_1 - E)}$ is $\gg 1$ Eq (31a) is closely approximated by

$$|T|^2 = \frac{16E(V_1 - E)}{V_1^2} \exp\left[-2a(2m|E - V_1|)^{1/2}/\hbar\right] \quad (31b)$$

where the exponential factor sometimes termed the barrier penetrability is $\ll 1$ The transmission through a less simple barrier $V(x)$ than Fig 4b is measured by the penetrability factor

$$P = \exp \frac{-2}{\hbar} \int_{x}^{x_2} dx\, (2m|E - V(x)|)^{1/2} \quad (32)$$

where x_1 x_2 are the turning points $E = V(x_1) = V(x_2)$ The barrier penetrability governs the rates at which (i) an incident proton whose kinetic energy is less than the height of the repulsive Coulomb potential barrier surrounding a nucleus is nonetheless able to penetrate the nucleus to produce nuclear reactions (ii) α particles can escape from a radioactive nucleus (iii) electrons can be pulled out of a metal by a strong electric field *See* FIELD EMISSION FUSION NUCLEAR RADIOACTIVITY



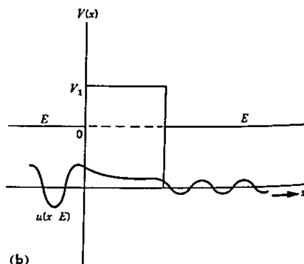


Fig 4 (a) Reflection from an infinitely thick barrier (b) Reflection from a barrier of finite thickness

Harmonic oscillator The potential $V(x) - \frac{1}{2}Kx^2$ Fig 5 describes a classical harmonic oscillator whose equilibrium position is $x - 0$ the classical oscillator frequency is $\nu = (2\pi)^{-1}\sqrt{K/m}$ Since $V(x)$ becomes infinite at $x = \pm\infty$ there is no continuous spectrum but there must be an infinite number of discrete eigenvalues Various sophisticated methods exist for finding the eigenfunctions and eigenvalues (*see* HARMONIC OSCILLATOR) The energy levels turn out to be

$$E_n = (n + \tfrac{1}{2})h\nu \quad (33)$$

where $n = 0\ 1\ 2$ (Fig 5) The corresponding first three eigenfunctions $u_0(\xi)$ u_1 u_2 are sketched in Fig 6 with $\xi - x(mK)^{1/4}/\hbar^{1/}$ a convenient dimensionless variable the dashed vertical lines indicate the turning points at $\xi - \pm\sqrt{2n+1}$ Figure 7 plots the probability density $|u(\xi)|^2$ for $n - 10$ The classical probability of finding a particle in the x interval dx is proportional to the time spent in dx The curved dashed line in Fig 7 plots the classical probability density for a classical oscillator whose energy is $E_{10} - 21\frac{1}{2}h\nu$ Eq (33)

The agreement between the classical probability density and the average over oscillations of $|u_{10}(\xi)|^2$ illustrates the connection between classical particle mechanics and the more fundamental dual wave particle interpretation of quantum theory With increasing n the oscillations of $|u_n(\xi)|^2$ become more rapid and the agreement with the classical probability density improves in accordance with the correspondence principle These harmonic oscillator results are a good first approximation to the energy levels and eigenfunctions of vibrating atoms in isolated molecules and in solids *See* LATTICE VIBRATIONS MOLECULAR STRUCTURE AND SPECTRA SPECIFIC HEAT OF SOLIDS

EXPECTATION VALUES

Suppose B is an operator that commutes with A If $Au_n - \alpha_n u_n$ $BAu_n = B(\alpha_n u_n)$ so that using Eq (6)

$$A(Bu_n) - \alpha_n(Bu_n) \quad (34)$$

Equation (34) means that Bu_n also is an eigenfunction of A corresponding to the eigenvalue α_n When α_n is not degenerate Eq (14) means Bu_n must be a multiple of u_n that is $Bu_n = \beta u_n$ β being a constant thus u_n is simultaneously an eigenfunction of the pair of commuting operators A B When the order of degeneracy of α_n is d Bu_n need not be a multiple of u_n but d independent eigenfunctions $u_{n1}(\alpha_n\ \beta_1)$ $u_{nd}(\alpha_n\ \beta_d)$ always can be found such that $u_n\ (\alpha_n\ \beta_s)$ is an eigenfunction of B corresponding to the eigenvalue [illegible] corresponding [illegible] u_{nd} be

values $\langle A\rangle$ of all operators A are time-independent … In … and … une- … dependent for example $A = A(\mathbf{r}\,\mathbf{p})$ but not $A(\mathbf{r}\,\mathbf{p}\,t)$ that

$$\frac{d}{dt}\langle A\rangle = \int \left[\psi^*\left(A\frac{\partial\psi}{\partial}\right) + \frac{\partial\psi^*}{\partial t}(A\psi)\right]$$
$$= \frac{1}{i\hbar}\int [\psi^*(AH\psi) - (H\psi)^*A\psi] \tag{49}$$
$$= \frac{1}{i\hbar}\int [\psi^*(AH - HA)\psi] = \frac{1}{i\hbar}\langle(AH - HA)\rangle$$

Equation (49) uses the notation of Eq (10) Of course $\langle(AH - HA)\rangle = 0$ whenever $\psi(t)$ is a stationary state $u(E\,t)$ Equation (49) shows however that if A commutes with the Hamiltonian then $\langle A\rangle$ is independent of time whether or not the system is in a stationary state Consequently operators commuting with the Hamiltonian are termed constants of the motion a system initially described by an eigenfunction $u(E\,\beta\,\gamma\ \ldots)$ of the simultaneously commuting observables $H\ B\ C\ \ldots$ remains in an eigenstate of $H\ B\ C\ \ldots$ as the wave function evolves in time

Equation (49) is closely analogous to the classical mechanics expression for $dA(\mathbf{r}\,\mathbf{p})/dt$ where $A(\mathbf{r}\,\mathbf{p})$ is the classical quantity corresponding to the quantum mechanical operator A If A is put equal to I in Eq (49) one obtains

$$\frac{d}{dt}\int \psi^*\psi = \frac{1}{i\hbar}\int [\psi^*(H\psi) - (H\psi)^*\psi] \tag{50}$$

Therefore the requirement that H be Hermitian Eq (8) which has been justified on the grounds that the eigenvalues of H must be real has the further important consequence that the right side of Eq (50) is zero, that is that Eq (2a) is obeyed at all times $t > t_0$ if it is obeyed at $t = t_0$ This re… It is necessary for the consistency of the formal… otherwise … could not be claimed that $|\psi(t)|^2$ from Eq (46) is the probability density at $t > t_0$ For a single spinless three-dimensional particle with $H = p^2/2m + V(x\,y\,z)$ it follows directly from Eq (46) that

$$\frac{\partial}{\partial t}(\psi^*\psi) + \frac{\partial}{\partial x}S_x + \frac{\partial}{\partial y}S_y + \frac{\partial}{\partial z}S_z = 0 \tag{51a}$$

where

$$S_x = \frac{\hbar}{2mi}\left(\psi^*\frac{\partial\psi}{\partial x} - \psi\frac{\partial\psi^*}{\partial x}\right) \tag{51b}$$

and so on In Eq (51a) $S_x\ S_y\ S_z$ can be interpreted as the components of a probability current vector $\mathbf{S}$ whose flow across any surface enclosing a volume τ accounts for the change in the probability of finding the particle inside τ See EQUATION OF CONTINUITY MAXWELL'S EQUATIONS

For a nonquadratically integrable ψ in the one particle case where Eq (51b) is applicable the probability current at infinity generally has the value $v\,|\psi|^2$ where v is the classical particle velocity at infinity the one-dimensional plane waves of … current density when $|\psi|^2$ is nonvanishing at infinity These considerations may be generalized to more complicated systems and are important in collision problems where the incoming and outgoing currents at infinity determine the cross section

Invariance Extremely general arguments support the view that the form of the Schrodinger equation (46) for any g particle system isolated from the rest of the universe must be (i) translation invariant that is, independent of the origin of coordinates (ii) rotation invariant that is independent of orientation of the coordinate axes and (iii) reflection invariant that is independent of whether one chooses to use a left handed or right handed coordinate system

The only known failures of these general requirements occur for reflections in a domain outside the scope of nonrelativistic quantum theory namely in phenomena such as beta decay that are connected with the so-called weak interactions Correspondingly it can be inferred that the Hamiltonian operator H for any such isolated system must commute with (i) the total linear momentum operator $\mathbf{p}_R = \mathbf{p}_1 + \ldots + \mathbf{p}_g$ (ii) the total angular momentum operator $\mathbf{J}$ (iii) the parity operator P which reflects every particle through the origin that is changes r_1 to $-r_1$ … r_g to $-r_g$ For additional information see PARITY (QUANTUM MECHANICS) SYMMETRY LAWS (PHYSICS) In quantum mechanics as in classical mechanics therefore linear momentum and total angular momentum are conserved that is are constants of the motion Since for an infinitesimal displacement ϵ in the x direction

$$\psi(x_1 + \epsilon\ y_1\ z_1\ x_2 + \epsilon\ y_2\ z_2\ \ldots\ x_g + \epsilon\ y_g\ z_g)$$
$$= \psi(x_1\,y_1\,z_1\ \ldots\ x_g\,y_g\,z_g) + \epsilon\left(\frac{\partial\psi}{\partial x_1} + \frac{\partial\psi}{\partial x_2} + \ldots + \frac{\partial\psi}{\partial x_g}\right)$$
$$= \psi + \epsilon p_{Rx}\psi \tag{52}$$

the connection between $p_{1x} + \ldots + p_{gx}$ and translation in the x direction can be understood the connection between $\mathbf{J}$ and rotation is under…

where the time-dependent potential energy $V(t)$ represents an externally imposed interaction for example with measuring equipment it is supposed that $V(t) = 0$ for $t < 0$ and $t > t_1$ Usually the system is in a stationary state $u(E_i)$ of H_0 at times $t < 0$ and one wishes to compute the probability of finding the system in some other stationary state $u(E_f)$ of H_0 at times $t > t_1$ From Eq (36) and the discussion preceding Eq (49)

in the notation of Eq (36) Equations (36) and (38) make it reasonable to postulate that whenever a system is instantaneously described by the normalized quadratically integrable $\psi(\mathbf{r}_1, \ldots, \mathbf{r}_g)$, the instantaneous probability of finding the system in the state corresponding to $A = \alpha$ is

$$P(\alpha) = \int d\beta \int d\gamma \cdots |c(\alpha,\beta,\gamma, \ldots)|^2 \quad (39a)$$

Similarly, the simultaneous probability of finding the system in the state corresponding to $A = \alpha$ and $B = \beta$ is

$$P(\alpha,\beta) = \int d\gamma \cdots |c(\alpha,\beta,\gamma, \ldots)|^2 \quad (39b)$$

and so on Clearly Eqs (1) and (2) are a special case of Eqs (38) and (39), in which the coordinate operators, $\mathrm{x}_1, \mathrm{y}_1, \mathrm{z}_1, \ldots, \mathrm{x}_g, \mathrm{y}_g, \mathrm{z}_g$ of the g spinless particles form the complete set of simultaneously commuting observables, and $\psi(x_1,y_1,z_1, \ldots, x_g, y_g, z_g)$ is the projection of ψ on the eigenfunction simultaneously corresponding to $\mathrm{x}_1 = x_1, \mathrm{y}_1 = y_1, \ldots, \mathrm{z}_g = z_g$

The total probability $\int d\alpha\, P(\alpha)$ is unity, Eq (38), and therefore any measurement of the observable A must yield a number equal to one of its eigenvalues Moreover, by the very meaning of probability, the average or expectation value of the observable A must be

$$\langle A \rangle = \int d\alpha\, \alpha\, P(\alpha) \quad (40a)$$

(see PROBABILITY) Equation (40a) shows that the same operator may have different forms in different representations For instance in the momentum representation discussed earlier where $\alpha \equiv k$, $p_x c(k)$ must equal simply $\hbar k c(k)$, consequently, to satisfy Eq (7), $xc(k)$ must equal $i(\partial c/\partial k)$ When ψ is a wave function for which Eq (8) holds with $\xi \equiv \psi$, Eqs (5), (12), (35), and (36) imply that

$$\langle A \rangle = \int \psi^* (A\psi) \quad (40b)$$

is equivalent to Eq (40a) Equation (40b), usually more convenient than Eq (40a), predicts the expectation value of any given observable in an arbitrary physical situation described by a reasonably well behaved quadratically integrable wave function, expectation values are not defined for non quadratically integrable ψ Equation (8) guarantees that $\langle A \rangle$ computed from Eq (40b) is a real number, necessary if $\langle A \rangle$ is to represent the results of measurement (see the earlier discussion on real eigenvalues)

Uncertainty principle. A precise measure of the spread or uncertainty in the value of A is ΔA defined by

$$(\Delta A)^2 = \langle (A - \langle A \rangle)^2 \rangle = \langle A^2 \rangle - (\langle A \rangle)^2 \quad (41)$$

Here $(\Delta A)^2$ is the average square deviation of A from its average $\langle A \rangle$ The quantity $\Delta A = 0$ when and only when ψ is an eigenfunction of A, that is, $A\psi = \alpha\psi$, in which event $\langle A \rangle = \alpha$ $\langle A^2 \rangle = \alpha^2$ The discussion following Eq (34) implies ΔA and ΔB simultaneously can be zero, that is A and B are simultaneously exactly measurable, whenever the commutator $AB - BA$ is zero If $AB - BA \neq 0$ introduce $A' = A - \langle A \rangle$, $B' = B - \langle B \rangle$, use Eqs (10) and (40b), and employ the so-called Schwarz inequality Then

$$(\Delta A)^2(\Delta B)^2 = \int (A'\psi)^*(A'\psi) \int (B'\psi)^*(B'\psi) \geq \left|\int (A'\psi)^*(B'\psi)\right|^2 = \left|\int \psi^*(A'B'\psi)\right|^2 \quad (42)$$

But

$$A'B' = \tfrac{1}{2}(A'B' - B'A') + \tfrac{1}{2}(A'B' + B'A') \quad (43)$$

which, after further manipulation, leads to

$$(\Delta A)^2(\Delta B)^2 \geq \tfrac{1}{4}|\langle AB - BA \rangle|^2 \quad (44)$$

Equation (44) is the rigorous quantum theoretic formulation of the uncertainty principle When $A = x$, $B = p_x$, Eq (7) yields

$$(\Delta x)(\Delta p_x) \geq \hbar/2 \quad (45)$$

This simple derivation of the uncertainty principle demonstrates anew the necessity for a dual wave-particle interpretation of the operator formalism

TIME DEPENDENCE

The procedures developed thus far predict the results of measurement at any given instant of time t_0 that the wave function $\psi(x\, t_0)$ is known Given ψ at t_0, to predict the results of measurement at some future instant t, it is necessary to know the time-evolution of ψ from t_0 to t It is postulated that this evolution is determined by the time-dependent Schrodinger equation

$$H\psi = i\hbar \frac{\partial \psi}{\partial t} \quad (46)$$

where H is the Hamiltonian or energy operator, ψ is supposed to be a continuous function of t If H were a number the solution to Eq (46) would be

$$\psi(t) = \exp\left[\frac{-iH(t - t_0)}{\hbar}\right]\psi(t_0) \quad (47)$$

Equation (47) remains valid when H is an operator, provided one understands that

$$\exp\left[\frac{-iH(t - t_0)}{\hbar}\right] \equiv I - \frac{i(t - t_0)}{\hbar} H + \frac{[-i(t - t_0)]^2}{2\hbar^2} H^2 + \cdots \quad (48)$$

with I the unit operator $I\psi = \psi$ the right side of Eq (48) is the usual series expansion of the exponential on the left side These so-called operational methods which manipulate operators like numbers, are widely employed in quantum theory, they must be used cautiously but usually lead rapidly to the same results as more conventional mathematical techniques

When $H\psi(t_0) = E\psi(t_0)$, that is when $\psi(t_0)$ is known to be an eigenfunction $u(E,t_0)$ of the energy operator, Eq (47) shows that $\psi(t) \equiv u(E,t) = u(E\, t_0) \exp[-iE(t - t_0)/\hbar]$ so that $|u(E\, t)|^2 = |u(E\, t_0)|^2$, similarly the expectation

$E = E_1 +$ $+ E_n$ Similarly if $H(x) - T_x + V(x)$ with eigenfunctions $u(x\,E_x)$ is a Hamiltonian for a one-dimensional spinless particle then any quadratically integrable $\psi(x\,y,z)$ describing a three dimensional spinless particle can be expanded in a series of products $u(x,E_x)\,u(y,E_y)\,u(z,E_z)$ This paragraph explains the relevance to three-dimen

ue o i u iiate ine gei eiai iormansm The reader is cautioned that the remaining contents of this article although very important especially for

parity operator has but two eigenvalues namely +1 and −1 the corresponding eigenfunctions are said to have even or odd parity Evidently P commutes with the harmonic oscillator Hamiltonian $p_x^2/2m + (1/2)Kx^2$ The harmonic oscillator eigenvalues are nondegenerate and therefore every harmonic oscillator eigenfunction (Fig 6) has either even or odd parity Similarly the eigenfunctions u' u of T_x Eq (21) have even and odd parity respectively eigenfunctions of T_x which do not have definite parity also exist however Eqs (23) because the eigenvalues E are degenerate The eigenfunctions of p_x do not have definite parity Eq (18a) because $p_xP + Pp_x = 0$ that is p does not commute but instead anticommutes, with P

Time evolution of packet A wave function ψ representing a single (spinless) particle localized in the neighborhood of a point is termed a wave packet Assuming $H = p^2/2m + V(\mathbf{r})$ Eq (49) yields

$$\frac{d}{dt}\langle x\rangle = \frac{1}{i\hbar}\langle xH - Hx\rangle = \left\langle\frac{p_x}{m}\right\rangle \quad (58a)$$

employing Eq (7) Similarly

$$\frac{d^2}{dt^2}\langle x\rangle = \frac{d}{dt}\left\langle\frac{p_x}{m}\right\rangle = \frac{-1}{m}\left\langle\frac{\partial V}{\partial x}\right\rangle \quad (58b)$$

Equations (58) mean (i) the average position of the particle that is, the center of the packet moves with a velocity given by the expectation value of the momentum (ii) the acceleration of the center of the packet is found from the expectation value of the classical force $-\partial V/\partial x$ Equations (58) illustrate the correspondence between quantum and classical mechanics and show that the classical description of particle motion is valid when the spread of the packet about its mean position can be ignored When the particle is free $H = p^2/2m$ Eqs (7) (41) and (49) lead to

$$(\Delta x)_t^2 = (\Delta x)_0^2 + \left\{\frac{t}{m}\langle xp + px\rangle_0 - 2\langle x\rangle_0\langle p\rangle_0\right\} + \frac{t^2}{m^2}(\Delta p_x)_0^2 \quad (59)$$

where the subscripts t 0 refer respectively to expectation values at t and at initial time zero Equation (59) shows that although an unconfined free wave packet may contract for a while ultimately it will spread over all space when the minimum spread happens to occur at $t - 0$ the term linear in t vanishes in Eq (59)

Orbital angular momentum The quantum mechanical operators representing the components of orbital angular momentum $\mathbf{L}$ have the same form as in classical mechanics namely for each particle $L_x = yp_z - zp_y$ and so on Using Eq (7)

$$L^2L_z - L_zL^2 = 0 \quad (60a)$$
$$L_xL_y - L_yL_x = i\hbar L_z \quad (60b)$$

and so on where $L^2 \equiv L_x^2 + L_y^2 + L_z^2$ According to Eq (44) therefore (i) L^2 and L_z are simultaneously exactly measurable (ii) once the values of L^2 and L_z are specified, the values of L_x and L_y must be uncertain In spherical coordinates $z = r\cos\theta$ $x = r\sin\theta\cos\phi$ $y = r\sin\theta\sin\phi$ L_z and L^2 become

$$L_z = \frac{\hbar}{i}\frac{\partial}{\partial\phi} \quad (61a)$$

$$L^2 = -\hbar^2\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right) \quad (61b)$$

Equation (12) for L_z is solved by $u(\phi\,m) = \exp(im\phi)$ where $m\hbar$ is the eigenvalue It can be argued that $u(\phi,m)$ must have a unique value at any point $x\,y\,z$ meaning $u(\phi + 2\pi\,m) = u(\phi\,m)$ so that m must be a positive or negative integer or zero With $\partial^2/\partial\phi^2 = -m^2$ in Eq (61b) the eigenvalues of L^2 turn out to be $l(l+1)\hbar^2$ where $l = 0$ 1 independent of m, except that for each l the allowed values of the magnetic quantum number m are $m = -l$ $-l+1$ $l-1$ l, thus each l has order of degeneracy $2l+1$ Because L^2 and L_z commute with P the eigenfunctions $u(l\,m)$ have definite parity in fact

$$Pu(l\,m) = (-1)^l u(l\,m)$$

In a two particle system the components of the total orbital angular momentum $\mathbf{L} = \mathbf{L}_1 + \mathbf{L}_2$ obey the same commutation Equations (60) ... ha ...

... and L_2^2 but L^2 does not commute with L_{1z} or L_{2z} Consequently the total orbital angular ...
labeled by
values of l ...
down to l =
tions have

These rules for combining angular momenta are readily generalized to more complicated systems including spin are well established and form the basis for the vector model of the atom See ATOMIC STRUCTURE AND SPECTRA.

Coulomb potential The Hamiltonian for an (assumed spinless) electron of mass m_e in the field of a fixed charge Ze is $H = \mathbf{p}^2/2m_e + V(r)$, where

$$\psi(t) = \int dE\, d\beta\, c(E\,\beta\,t) \exp(-iEt/\hbar)\, u(E\,\beta) \quad (53)$$

which when substituted in Eq (46) for times $0 \leqq t \leqq t_1$ yields

$$i\hbar \int dE\, d\beta \frac{dc(E\,\beta\,t)}{dt} \exp(-iEt/\hbar)\, u(E,\beta)$$
$$= \int dE\, d\beta\, c(E\,\beta\,t) \exp(-iEt/\hbar)\, V(t) u(E,\beta) \quad (54)$$

In Eqs (53) and (54) E is α of Eq (36) and β stands for all other indices β γ necessary to make $u(E)$ a complete orthonormal set if $V(t)$ were zero the projections $c(E\,\beta)$ would equal $\delta(E - E)\ \delta(\beta - \beta_i)$ independent of time

Most problems in quantum theory cannot be solved exactly therefore some approximate treatment using perturbations must be devised In the present case it is assumed that $c(E\,\beta\,t)$ do not change appreciably from their initial values at $t = 0$ so that it is a reasonable approximation to replace $c(E\,\beta\,t)$ by $c(E\,\beta\,0)$ on the right side of Eq (54) With the further approximation that V is constant during the interval $0 \leqq t \leqq t_1$ one finds using the notation of Eq (10) that

$$|c(E_f\,\beta_f\,t_1)|^2 - \frac{4}{\hbar^2}\left|\int u^*(E_f\,\beta_f)V\ u(E\ \beta)\right|^2 \frac{\sin^2 \frac{1}{2}\omega t_1}{\omega^2}$$
$$= \frac{4}{\hbar^2}|V_f|^2 \frac{\sin^2 \frac{1}{2}\omega t_1}{\omega^2} \quad (55)$$

where $\hbar\omega = E_f - E$

Equation (55) shows that the probability of finding the system in some new stationary state $u(E_f\,\beta_f)$ after the system is no longer perturbed is proportional to (i) $|V_f|$ the square of the matrix element of the perturbation between initial and final states (ii) an oscillating factor which for given t_1 has a peak $t_1/4$ at $\omega - 0$

$\sin^2(\frac{1}{2}\omega t_1)/\omega$ is plotted as a function of ω in Fig 8 evidently $|c(E_f\,\beta_f)|^2$ is relatively small for energies such that $|\omega| > \sim\pi/t_1$ The most likely occupied states after the perturbation conserve energy ($E_f = E$) and the spread in energy of the final states is $\Delta E - \hbar\,\Delta\omega = \sim 2\pi\hbar/t_1 \equiv h/\Delta t$ where $t_1 \equiv \Delta t$ is for example the duration of the measurement Thus Eq (55) provides a version of the uncertainty principle between energy and time As t_1 approaches infinity the area under the curve in Fig 8 becomes proportional to t_1 since the main peak has a height proportional to t_1^2 and a width proportional to $1/t_1$ Therewith one obtains a widely employed formula giving the approximate transition probability w per unit time for making transitions under the influence of a steady perturbation V from an initial stationary $u(E_i\,\beta_i)$ to a set of final states of the same energy namely

$$w = \frac{2\pi}{\hbar}\rho(E_f)|V_f|^2 \quad (56)$$

where $\rho(E_f) - d\beta\, d\gamma$ is the density of independent final states in the neighborhood of $E = E_f = E$ $\beta = \beta_f$ $\gamma = \gamma_f$ and so on For instance with u_i a plane wave e^{ikz} moving in the z direction Eq (18a) and u_f a plane wave in some other direction Eq (56) yields the Born approximation to the cross section for elastic scattering by a potential Equation (56) is also applicable to problems outside the domain of nonrelativistic quantum theory for example to the theory of β-decay wherein new particles are created

The preceding considerations are important for understanding how a measurement of an operator A not commuting with the Hamiltonian H can cause an initially stationary state $u(E)$ to evolve into an eigenstate $u(\alpha)$ of A Equations (46) (49) with $H = T + V = H_0$ the unperturbed Hamiltonian hold only in the intervals between measurements during the measurement $u(E)$ though an eigenfunction of H_0 is not a stationary state of the complete Hamiltonian This paragraph does not do justice however to the subtle questions involved in the quantum theory of measurement

FURTHER ILLUSTRATIVE APPLICATIONS

When g particles are noninteracting the Hamiltonian has the trivially separable form $H = H_1 + H_2 + \cdots + H_g$ and

$$(H_1 + \cdots + H_g)[u_1(\mathbf{r}_1\,E_1\,\alpha_1)\cdots u_g(\mathbf{r}_g,E_g\,\alpha_g)]$$
$$-(E_1 + \cdots + E_g)$$
$$[u_1(\mathbf{r}_1,E_1\,\alpha_1)\cdots u_g(\mathbf{r}_g,E_g\,\alpha_g)] \quad (57)$$

where $u_1(\mathbf{r}_1\,E_1\,\alpha_1)$ are a complete orthonormal set of eigenfunctions of H_1 for particle 1 and so on Thus because an operation performed solely on particle 1 always commutes with an operation performed solely on particle 2 the products $u_1(\mathbf{r}_1,E_1\,\alpha_1)\cdots u_g(\mathbf{r}_g,E_g\,\alpha_g)$ are a complete orthonormal set of eigenfunctions of $H = H_1 + \cdots + H_g$ also the energy levels of H are the set of possible sums of individual particle energies,

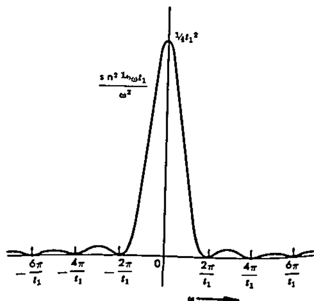


Fig 8 Plot of $\frac{\sin^2 \frac{1}{2}\omega t_1}{\omega^2}$ vs $\omega = (E_f - E_i)/\hbar$ (L. I Schiff Quantum Mechanics 2d ed McGraw Hill 1955)

trary spherically symmetric $V(r)$ the bound state eigenvalues $E(n,l)$ of Eq (62b) do not coincide for different l and each bound state has degeneracy $2l+1$ corresponding to the $2l+1$ possible values $m = -l$ to l of the magnetic quantum number m an energy level associated with orbital angular momentum l has parity $(-)^l$ For any such $V(r)$ the bound state energies (i) increase with increasing n for a constant value of l for the same reasons that were discussed in connection with the eigenvalues of Eq (25) (ii) increase with increasing l for constant n because the rotational kinetic energy $(2mr^2)^{-1}l(l+1)\hbar^2$ increases Except for these regularities the order and spacing of the levels depends on the details of $V(r)$ For potentials $V(r)$ decreasing more rapidly than $1/r$ the total number of bound states generally is finite whereas this number is infinite for the Coulomb $V(r) = -Ze^2/r$

Removal of degeneracy The Hamiltonian of Eq (62a) is spin independent which is why Eq (62a) has been tre ted

solution u of Eq (62a) there are two independent two-component eigenfunctions (i) $\psi_1 = u$ $\psi_2 = 0$ and (ii) $\psi_1 = 0$ $\psi_2 = u$ compare the earlier discussion of spin Thus for an electron the degeneracy of the energy levels in a Coulomb field is $2n$ similarly the degeneracy for neutrons protons or electrons in an arbitrary spherically symmetric potential is $2(2l+1)$ The energy operator for an electron in an actual atom for example hydrogen is not spin independent however Relativistic effects add to the central $V(r)$ of Eq (62a) noncentral spin orbit potentials $V'(r)[L_x s_x + L_y s_y + L_z s_z] \equiv V'(r)\mathbf{L}\cdot\mathbf{s}$ here $\mathbf{s}$ is the spin operator and obeys the same commutation equations (60) as $\mathbf{L}$

Equations (60) show that $V'(r)\mathbf{L}\cdot\mathbf{s}$ commutes with L^2 s^2 $(\mathbf{L}+\mathbf{s})^2$ and $L_z + s_z$ but not with L_z or s_z illustrating the principle of conservation of total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{s}$ compare the remarks preceding Eq (52) Consequently referring to the final part of the discussion of orbital angular momentum (i) $J^2 = j(j+1)\hbar^2$ (ii) for given $l \neq 0$ (and $s = \frac{1}{2}$) j has but two possible values $l \pm \frac{1}{2}$ (iii) the energy levels are labeled by j and have a $(2j+1)$ fold degeneracy corresponding to the $2j+1$ possible orientations of $\mathbf{J}$ $-j$ to $+j$ (iv) because $2\mathbf{L}\cdot\mathbf{s} = (\mathbf{L}+\mathbf{s})^2 - L^2 - s^2$ levels of different j have different energ and the an bly on l of the m still are oreover because $2j+1 = 2l+2$ for $j = l + \frac{1}{2}$ and $= 2l$ for $j = l - \frac{1}{2}$ the total number of independent eigenfunctions associated with given l (and n) remains $2(2l+1) = (2l+2) + 2l$

In the independent particle model of atoms and nuclei one assumes that to a first approximation each particle particle i say moves in a potential $V(r_i)$ which has been averaged over the coordinates of all the other particles so that $V(r_i)$ depends only on the distance of i from the atomic or nuclear center To a first approximation therefore the energy levels are associated with configurations of one particle eigenfunctions for example the ground state of atomic Be is $1s^2 2s^2$ In higher approximation one introduces two-body interactions $V(r_i r_j)$ which may be said to mix different configurations The considerations of this and the paragraphs just preceding together with the exclusion principle discussed subsequently account for the periodic system of the elements and are the basis for the highly successful nuclear shell model

The observation that splitting the levels does not change the number of independent eigenfunctions illustrates a general principle and justifies the postulate that the statistical weight of a discrete level equals its order of degeneracy (see BOLTZMANN STATISTICS STATISTICAL MECHANICS) This principle can be understood on the basis that the number of bound state eigenfunctions should be a continuous function of the parameters in a reasonably well behaved Hamiltonian because this number by definition is an integer it must change discontinuously and therefore except under unusual mathematical circumstances cannot change at all

In an external magnetic field B the Hamiltonian

about t e magnetic held so that (with the

magnetic quantum numbers $-j$ to $+j$ This prediction is thoroughly confirmed in the Zeeman effect and in the Stern Gerlach experiment See ZEEMAN EFFECT

Radiation The classical Hamiltonian for a charged particle in an electromagnetic field has the form

$$H = \frac{1}{2m}\left(\mathbf{p} - \frac{e\mathbf{A}}{c}\right)^2 + e\phi \qquad (65)$$

where $\mathbf{A}$ and ϕ are the scalar and vector potentials respectively (see ELECTRON MOTION IN VACUUM) It is postulated that when properly symmetrized that is, when $\mathbf{A}\cdot\mathbf{p} + \mathbf{p}\cdot\mathbf{A}$ replaces the classical $2\mathbf{A}\cdot\mathbf{p}$ (see the earlier discussion on Hermitian operators) H of Eq (65) is the quantum mechanical energy operator The presence of terms linear in $\mathbf{p}$ modifies some of the formulas which have been given for example Eq (51a) When plane waves of light (frequency f wavelength λ and moving in the z direction) fall on a hydrogen atom $\mathbf{A}$ is proportional to $\cos[2\pi(z/\lambda - ft)]$ and $e\phi$ is the Coulomb potential $V(r)$ of Eq (62a)

Proceeding as in Eqs (53)–(56) noting that $\mathbf{A}$ contains terms proportional to $\exp(2\pi ift)$ and

$V(r) = -Ze/r$ In spherical coordinates Eq (12) for the eigenfunctions $u(r\,\theta\,\phi)$ is

$$Hu = \left[\frac{-\hbar^2}{2m_e}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{1}{2m_e r^2}L^2 + V(r)\right]u = Eu \qquad (62a)$$

with L^2 defined by Eq (61b) Now H commutes with L^2 and L_z and $r\,H$ is separable that is

$$r^2H(r\,\theta\,\phi) = H_1(r) + (2m_e)^{-1}L^2(\theta\,\phi)$$

[compare Eq (57)] Thus $u(r)u(l\,m)$ are a complete set of eigenfunctions $L^2u(r)u(l\,m) = l(l+1)\hbar^2u(r)u(l\,m)$ and therefore the radial eigenfunctions $u(r) \equiv u(E\,l)$ must satisfy

$$\left[\frac{-\hbar^2}{2m_e}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right) + \frac{l(l+1)\hbar^2}{2m_e r^2} + V(r)\right]u(r) = Eu(r) \qquad (62b)$$

The positive term $l(l+1)\hbar^2/2m_e r^2$ acts as an added repulsive potential it can be understood in terms of the classical centripetal force needed to maintain the angular momentum For $E < 0$ admissible solutions to Eq (62b) must be exponentially decreasing at $r = \infty$ moreover because of the r^{-1} and r^{-2} terms in Eq (62b) an eigenfunction $u(E\,l)$ which behaves properly at $r = \infty$ becomes infinite at $r = 0$ unless E is specially chosen Thus (as always) the quadratically integrable eigenfunctions form a discrete set The corresponding bound state energies $E < 0$ are

$$E = -\frac{m_e Z^2 e^4}{2\hbar^2 n^2} \qquad (63)$$

In Eq (62) the principal quantum number $n = 1\ 2\ 3$ for given l and n the number $n' \geq 0$ of zeros (between $r = 0$ and $r = \infty$) of the corresponding $u(E\,l)$ is $n' = n - l - 1$ Because $dx\,dy\,dz = r^2 \sin\theta\,dr\,d\theta\,d\phi$ the radial probability density is $r^2|u(r\,E\,l)|^2$ Figure 9 shows the radial probability density plotted vs r (in units of the Bohr radius $a_0 = \hbar^2/m\,e^2 \cong 5 \times 10^{-9}$ cm) for several low lying stationary states of atomic hydrogen $Z = 1$ The notation for the eigenfunctions is standard in atomic physics the principal quantum number is supplemented by a lower case letter $s\ p\ d$ corresponding to $l = 0\ 1\ 2$ for example a $3d$ state has $l = 2$ and therefore $n' = 0$ or no radial nodes as in Fig 9 The eigenfunction $u(l = 0\ m = 0)$ is a constant that is an s state is spherically symmetric $|u(r\,\theta\,\phi)|^2$ is proportional to $\cos^2\theta$ or to $\sin^2\theta$ in p states and so on The eigenfunctions although they are spread over all space have their maxima at about the radii expected on the older Bohr theory of Eq (63)

In the actual hydrogen atom the nucleus of mass M is not fixed The Hamiltonian is

$$H = \frac{p_1^2}{2m_e} + \frac{p_2^2}{2M} - \frac{Ze^2}{|r_1 - r_2|} \qquad (64a)$$

where the subscripts 1 and 2 refer to the electron and the nucleus respectively Introducing the center of mass $(X\,Y\,Z) \equiv \mathbf{R} = (M + m_e)^{-1}[m_e\mathbf{r}_1 + M\mathbf{r}_2]$ $(x\,y\,z) \equiv \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ Eq (64a) takes the separable form

$$H = \frac{p_R^2}{2(M + m_e)} + \frac{p^2}{2\mu} - \frac{Ze^2}{r} \qquad (64b)$$

where $(\mathbf{p}_R)_x \equiv p_X = (\hbar/i)\partial/\partial X$, etc are the components of the total momentum $\mathbf{p}_R = \mathbf{p}_1 + \mathbf{p}_2$ $p_x = (\hbar/i)\partial/\partial x$ etc and the reduced mass $\mu = m_e M(M + m_e)^{-1}$ Therefore $\mathbf{p}_R$ is a constant of the motion as asserted in connection with Eq (52) moreover Eq (64b) is separable in $\mathbf{R}$ and $\mathbf{r}$ In other words the center of mass moves like a free particle completely independent of the internal motion (*see* CENTER OF MASS) Comparing Eqs (62a) and (64b) and recalling Eq (57) the eigenvalues of Eq (64a) after the kinetic energy of the center of mass is subtracted are given by Eq (63) with μ (depending on m_e/M) replacing m_e The independence of internal and center of mass motion means that the temperature broadening of spectral lines can be explained quantum mechanically in terms of the classical Doppler effect for a moving fixed frequency source This paragraph illustrates the correspondence between the classical and quantum theories

The eigenvalues E of Eq (63) have order of degeneracy n^2 this degeneracy stems from (i) the fact that $V(r)$ is spherically symmetric which permits H to commute with $\mathbf{L}$ (ii) a special symmetry of Eq (62a) for the specially simple Coulomb potential causing solutions of Eq (62b) for different l to have the same energy For an arbi

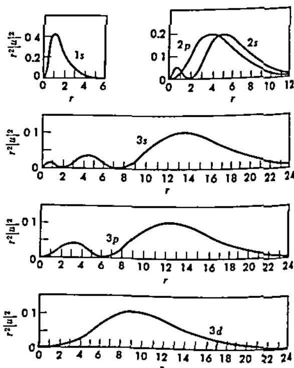


Fig 9 Radial probability density in atomic hydrogen (F K Richtmyer E H Kennard and T Lauritsen Introduction to Modern Physics 5th ed McGraw Hill 1955)

the many electron Schrödinger equation must be employed Similarly since a deuteron is a neutron proton bound state with total angular momentum unity in a deuterium molecule (i) each deuteron can be treated as if it were a fundamental particle of spin 1 (ii) the wave function of a deuterium molecule must be symmetric under interchange of the space and spin coordinates of the two deuterons which interchange involves successive antisymmetric interchanges of the two neutrons and of the two protons In other words (when they can be treated as particles) deuterons and other composite systems of integral spin obey Bose Einstein statistics, composite systems of half integral spin obey Fermi Dirac statistics

When the particles composing a many particle system can be represented by nonoverlapping wave packets which spread an amount $\ll \Delta x$ as the center packet moves a distance equal to its width Δx individual classical particle trajectories can be distinguished under these circumstances therefore the particles who be [illegible]

e classical Maxwell Boltzmann statistics The ell known condition for the validity of classical atistics in an electron gas $Nh^3(2\pi mkT)^{-3/2} \ll 1$ nplies that such packets can be constructed here is the electron density k is Boltzmann s constant is the absolute temperature In the lower vibrational states of a molecule such packets cannot be constructed for the vibrating nuclei however (compare the earlier discussion of the harmonic oscillator) so that quantum statistics cannot be ignored for example in the specific heat of H_2 at low temperatures *See* QUANTUM STATISTICS [E G]

Bibliography P A M Dirac *Principles of Quantum Mechanics* 4th ed 1958 H A Kramers, *Quantum Mechanics* 1957 L D Landau and E M Lifshitz *Quantum Mechanics Non relativistic Theory* 1958 L Pauling and E B Wilson Jr *Introduction to Quantum Mechanics* 1935, L I Schiff *Quantum Mechanics* 2d ed 1955 *see also* QUANTUM MECHANICS.

Quantum theory, relativistic

The form of quantum mechanics applicable to a system containing a fixed number of unchangeable particles whose velocities are equal or nearly equal to the velocity of light when particle creation or annihilation can occur quantum field theory must be employed (*see* QUANTUM FIELD THEORY) In practice relativistic quantum theory (as defined here) is limited to single-particle systems there is not yet a well-established relativistic generalization of the nonrelativistic many particle Schrodinger equation This article presupposes familiarity with the contents of three other articles (*see* QUANTUM MECHANICS QUANTUM THEORY NONRELATIVISTIC RELATIVITY)

Invariance It is a fundamental tenet that the formal structure of nonrelativistic quantum mechanics for example the Schrodinger equation must be invariant with respect to rotations and reflections of the coordinate axes Acceptance of the *special theory of* relativity implies that quantum mechanics must be invariant with respect to Lorentz transformations which generalize and include spatial rotations and reflections Thus nonrelativistic quantum theory is a first approximation *to a more* rigorously correct Lorentz invariant theory whose departures *from the simpler nonrelativistic* theory are most important when particle velocities approach the velocity of light *See* LORENTZ TRANSFORMATIONS SYMMETRY LAWS (PHYSICS)

In relativistic (as in nonrelativistic) quantum *theory the* formalism must embody wave particle duality, for example, the basic wave property of superposition, *it is also desirable to assume that* the wave function $\psi(t)$ at any time $t > 0$ is determined *solely by the values of* ψ *at* $t = 0$ The very successful nonrelativistic Schrodinger *equation*

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi \qquad (1)$$

has *the simplest* (very nearly the only) form consistent with these requirements Consequently one postulates that the relativistic wave equation also has the form (1) With the relativistic Hamiltonian *operator* H this postulate preserves the correspondence between classical and quantum mechanical *motion in the limit* $\hbar \to 0$ Unfortunately with $p_x = (\hbar/i)\,(\partial/\partial x)$ etc inserting the usual *relativistic free-particle Hamiltonian* $H = \sqrt{p^2c^2 + m^2c^4}$ into Eq (1) yields an equation which is unsatisfac*tory* on several counts for example *if the square* root is interpreted by its *series expansion* in powers *of* p the equation involves spatial derivatives of infinitely high order

Dirac equation *If it is assumed that classical* relativistic mechanics does not *exhaust* the pos*sible quantum* mechanical energy *operators* the simplest free particle *Lorentz* invariant equation of the form (1) is

$$i\hbar \frac{\partial \psi}{\partial t} = [-c\alpha \cdot \mathbf{p} - \beta mc^2]\psi$$
$$= \left[i\hbar c\left(\alpha_x \frac{\partial}{\partial x} + \alpha_y \frac{\partial}{\partial y} + \alpha_z \frac{\partial}{\partial z}\right) - \beta mc^2\right]\psi \qquad (2)$$

(*see* RELATIVISTIC MECHANICS) *In Eq* (2) *the* celebrated Dirac equation (P A M Dirac 1928) ψ is a four component wave function *meaning* (in nonrelativistic language) that the particle has spin $\frac{1}{2}$ and $\alpha \equiv \alpha_x\ \alpha_y\ \alpha_z$ are spin-dependent *but* coordinate independent operators represented by four rowed square matrices With $i\hbar\,\partial\psi/\partial t = E\psi$ *there* is a continuum of eigenfunctions of *Eq* (2) corresponding to all positive energies $E \geq mc^2$ When E approaches the particle rest energy mc^2 that is when the particle velocity approaches zero these eigenfunctions of *Eq* (2) have a pair of vanishing components Since only two *of the* four components are appreciable *in the nonrelativistic limit, it appears*

$\exp(-2\pi i f t)$, and neglecting the small (as can be shown) A^2 terms, one obtains an expression similar to Eq (55), except that $\omega \pm 2\pi f$ replaces $\omega \equiv \hbar^{-1}(E_f - E_i)$ In other words, after a long time there are appreciable transition probabilities only to final states f whose energies satisfy $E_f - E_i \pm hf = 0$, in agreement with the notion that a quantum of energy hf has been emitted or absorbed, the + sign corresponds to emission, the − sign to absorption The corresponding transition probabilities are given by Eq (56) with V' proportional to $\exp[2\pi i z/\lambda]$, and u_i, u_f stationary state atomic wave functions satisfying the radiation unperturbed Eq (62*a*) The final expressions are analogous to the classical formulas for emission or absorption of radiation, for instance, when the wavelength λ is large compared to atomic dimensions expanding $\exp[2\pi i z/\lambda]$ in powers of z/λ shows that the leading term in the transition probability is the matrix element, between initial and final states, of the dipole moment ez corresponding to classical electric dipole emission or absorption (*see* ELECTROMAGNETIC RADIATION) Because z changes sign on reflection through the origin the dipole matrix element vanishes unless u_i and u_f have opposite parities This is one of the selection rules for electric dipole radiation, other selection rules, connected with angular momentum conservation, are obtained similarly *See* SELECTION RULES (PHYSICS)

The theory starting with Eq (65) is termed semiclassical, because it does not replace the classical A, ϕ by a quantum mechanical operator description of the electromagnetic field This semiclassical theory has led to the induced emission and absorption probabilities in the presence of external radiation but not to the spontaneous probability of emission of a photon in the absence of external radiation The spontaneous transition probabilities can be inferred from the induced probabilities by thermodynamic arguments The spontaneous transition probability is deduced directly, however, without appeal to the arguments of thermodynamics, when the radiation field is quantized

PARTICLE INDISTINGUISHABILITY

For systems of g identical and therefore indistinguishable particles the formalism is further complicated because the probability P of finding a given particle in a specified volume $dx\,dy\,dz$ of space must be $P_1 + P_2 + \cdots + P_g$ where $P_1, \ldots, P_g$ are the probabilities given previously for distinguishable particles, expectation values and normalizations must be reinterpreted accordingly Moreover the Pauli exclusion principle asserts that the only physically permissible wave functions must change sign when the space and spin coordinates of any pair of indistinguishable particles of spin ½ are interchanged (*see* EXCLUSION PRINCIPLE) To amplify this assertion consider the four component wave function of the two electrons in atomic helium

$$\psi = \psi_{++}(\mathbf{r}_1,\mathbf{r}_2),\ \psi_{+-}(\mathbf{r}_1,\mathbf{r}_2),\ \psi_{-+}(\mathbf{r}_1,\mathbf{r}_2),\ \psi_{--}(\mathbf{r}_1,\mathbf{r}_2)$$

$|\psi_{++}(\mathbf{r}_1,\mathbf{r}_2)|^2$ is the probability density for finding both electrons with spin along $+z$, and so forth. The exclusion principle requires

$$\begin{aligned}\psi_{++}(\mathbf{r}_1,\mathbf{r}_2) &= -\psi_{++}(\mathbf{r}_2,\mathbf{r}_1)\\ \psi_{--}(\mathbf{r}_1,\mathbf{r}_2) &= -\psi_{--}(\mathbf{r}_2,\mathbf{r}_1)\\ \psi_{+-}(\mathbf{r}_1,\mathbf{r}_2) &= -\psi_{-+}(\mathbf{r}_2,\mathbf{r}_1)\end{aligned}$$

In the independent particle approximation all components of the ground state eigenfunctions of atomic He are composed of products $u(\mathbf{r}_1)u(\mathbf{r}_2)$ where $u(\mathbf{r})$ is the lowest 1*s* eigenfunction solving Eq (62*a*) for the spherically symmetric $V(r)$ appropriate to He In this approximation therefore, $\psi_{++} = \psi_{--} = 0$, if $\psi_{+-} = u(\mathbf{r}_1)u(\mathbf{r}_2)$, ψ_{-+} is necessarily equal to $-u(\mathbf{r}_1)u(\mathbf{r}_2)$ Of the four independent $1s^2$ eigenfunctions originally possible only one remains, which can be shown to be a total spin zero eigenfunction, the exclusion principle literally has excluded the three eigenfunctions corresponding to total spin one

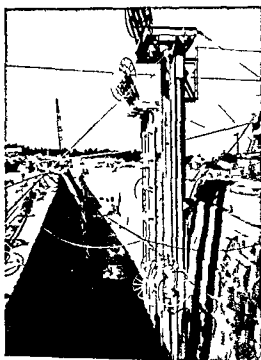
These results are summarized by the rule that at most two electrons $s_z = \pm\frac{1}{2}$, can occupy one-particle states with the same quantum numbers n_r, l, and $m = -l$ to l By the general principle explained previously in the discussion of removal of degeneracy, the introduction of two particle interactions $V(\mathbf{r}_1,\mathbf{r}_2)$ does not change the number of independent eigenfunctions consistent with the exclusion principle In the next higher 1*s*2*s* configuration of He ψ_{+-} will equal $u_{1s}(\mathbf{r}_1)u_{2s}(\mathbf{r}_2) - u_{1s}(\mathbf{r}_2)u_{2s}(\mathbf{r}_1)$ This antisymmetrized wave function makes nonclassical exchange energy contributions, of the form

$$\int u_{1s}{}^*(\mathbf{r}_1)u_{2s}{}^*(\mathbf{r}_2)V(\mathbf{r}_1,\mathbf{r}_2)u_{1s}(\mathbf{r}_2)u_{2s}(\mathbf{r}_1)$$

to the expectation value $\langle V(\mathbf{r}_1\,\mathbf{r}_2)\rangle$ of the interaction energy Exchange energies are important for understanding chemical binding

Except for the complication of spin, this article presupposes that electrons, neutrons, and protons are immutable structureless mass points Actually modern theories of nuclear forces and high energy scattering experiments indicate that this assumption is untrue When creation destruction, or other alterations of fundamental particle structure are improbable, however the general separability of internal and center of mass motion [see the discussion following Eq (64*b*)] implies that each fundamental particle is sufficiently described by the position of its center of mass and by its spin orientation that is by the many component wave functions used here

In circumstances wherein more obviously composite systems undergo no changes in internal structure, they too can be treated as particles For instance, in the slow collisions of two atoms the slowly changing potentials acting on the electrons do not induce transitions to new configurations and the collision can be described by solving an equation of the form (62*a*) for the relative motion, in rapid collisions electron transitions occur, and



New England gran te quarry Four parallel wire saws are cutt ng blocks 36 in apart the four w res appear emerg ng from the r cuts at the lower r ght (H E Fletcher Gran te Co)

ornamental use Hence selection of a suitable depos t is of primary importance The major problem in dimension stone quarrying is to secure large sound and relatively flawless blocks of stone of at tract

it on a track and cut a narrow channel by a chopping action The channels may be 10 or 12 ft deep and 4 ft apart A modern method of making initial cuts for primary block separations is by use of the wire saw which is simply a wire belt to which sand or other al rasive is fed with a stream of wa ter When th s l elt traveling under tension is brought in contact with the rock it cuts a channel about ¼ in wide This equipment is now used suc cessfully in Pennsylvania slate quarries in a New England granite quarry and in some Indiana lime-stone operations Another method of making pri mary cuts is known as drilling and broa h

If no horizontal open bed seams are present in a rock formation the masses of stone separated by channeling or otherwise are set free at the quarry floor by driving wedges into horizontal drill holes The large masses of stone bdivided into smaller blocks usually by the plug and feather method The feathers are elongated soft iron plates used in pairs (one down each side of a drill hole) and the plug is a steel wedge placed between a pair of feathers When they are inserted in a row of shal low drill holes and sledged lightly in succession a fracture is made

In some slates and many of the harder rocks such as granites and quartzites the primary breaks are usually made by discharging light powder charges in deep drill holes The charges are small enough to fracture without shattering the stone

The rough blocks are hoisted from the quarry with powerful derricks and are fashioned into fin ished products by sawing planing rubbing and polishing in mills equipped with a great variety of stone working machines

Crushed stone Quarrying to obtain crushed stone is an entirely different process Rows of holes are sunk even as much as 50 or 100 feet with per cussion or rotary drills to the bottom of a retreat ing bench within a quarry Charges some reaching several tons of explosives are fired simultaneously in these holes Thus masses of stone possibly thou sands of cubic yards are shattered and thrown to the quarry floor ready for loading This is known as the primary blast If some masses of stone are too large for loading they may be reduced in size by secondary blasting with small charges of dyna

the crushing operation The stone is loaded into dump cars or trucks with which it is hauled to the primary crusher usually of the jaw or gyratory type Such crushers reduce the stone to about 6 in size

The finer materials from the crusher discharge are removed with a screen usually of the rotary type Stone which will not pass through the screen is carried to smaller secondary crushers that reduce it to 1 or 1½-in size The crushed stone is graded by screening to various sizes depending upon the specifications of the users Most screens are of the shaking or vibrating type

Limestone is the rock most widely used in crush ing but granite and basaltic rock are also extensively ties. Crus aggregate ularly lim many other applications Great tonnages of limestone are available in the tailings from the lead and zinc mining operations in Missouri *See* MINING OPEN CUT OR PIT STONE AND STONE PRODUCTS

[O B]

Bibliography O Bowles *The Stone Industries* 2d ed 1939

that Eq (2) despite its four components corresponds to a particle of spin $s = \frac{1}{2}\hbar$ a conclusion which can be demonstrated rigorously from the transformation properties of ψ under spatial rotations *See* SPINOR

When a spherically symmetric potential operator $V(r)$ is added to the Dirac Hamiltonian and the small components are eliminated the resultant two component equation includes a spin orbit interaction needed to understand the level splitting in atomic spectra With $V = -e/r$ and m equal to the electron mass the energy levels coincide with the observed levels of atomic hydrogen except for very small quantum electrodynamic and nuclear corrections In an external magnetic field **B** where (as in the nonrelativistic case) **p** is replaced by $\mathbf{p} - e\mathbf{A}/c$ in the Dirac Hamiltonian the resultant two component equation contains an added interaction $-(e/mc)\mathbf{s}\cdot\mathbf{B}$ Thus the Dirac equation predicts the electron has a magnetic moment of magnitude (e/mc) $(\hbar/2)$ a prediction in accord with a variety of observations For these reasons it is considered established that the Dirac equation correctly describes the motion of an electron Protons and neutrons which also have spin $\hbar/2$ apparently are not described by the Dirac equation in its simple form (2) since the Dirac equation does not predict correctly the magnetic moments of these particles the μ meson whose magnetic moment seems to be predicted correctly presumably does obey the Dirac equation *See* ELEMENTARY PARTICLE MESON

Hole theory of positron The free particle Dirac equation (2) also has a continuous set of negative energy levels $E \leq -mc$ which somehow must be inaccessible otherwise all electrons would continually be making transitions to negative energy states and thence to states of even more negative energy etc Since electrons obey the Pauli exclusion principle Dirac suggested that these infinitely many negative energy levels normally are inaccessible to laboratory electrons because they already are occupied by other unnoticed electrons Absorption of a photon of energy $hf > 2mc^2$ can knock a negative energy electron into a positive energy state leaving behind a hole in the sea of occupied negative energy states It can be shown that this hole moves as if it were a positively charged electron Thus taken together with the exclusion principle the Dirac equation predicts the existence of the positron and the phenomenon of pair production similarly an electron can fall into a hole observed in the laboratory as annihilation of a positron electron pair In quantum field theory a hole theory of the positron can be avoided however *See* PAIR PRODUCTION (ELECTRON POSITRON) POSITRON POSITRONIUM

***Klein Gordon* equation** Whereas for a relativistic electron all four components are needed it appears (1959) possible to describe a neutrino whose spin also is $\frac{1}{2}\hbar$ but whose (rest) mass is zero and velocity always is c by means of only two components *See* NEUTRINO PARITY (QUANTUM MECHANICS) Other Lorentz invariant equations have been proposed in particular the Klein Gordon equation

$$-\hbar^2c^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\psi = m^2c^4\psi \quad (3)$$

obtained from $E^2 = p^2c^2 + m^2c^4$ using $E \to i\hbar\,\partial/\partial t$ $p_x \equiv (\hbar/i)(\partial/\partial x)$ etc in fact Eq (2) implies Eq (3) Using Eq (3) alone for an electron ignoring the subsidiary conditions implied by Eq (2) leads to difficulties connected with the fact that Eq (2) does not have the desirable form (1) Although Eq (3) is acceptably Lorentz invariant for particles of arbitrary spin as yet it is not known whether any particles obey Eq (3) without subsidiary conditions *See* QUANTUM ELECTRODYNAMICS [E.G.]

Bibliography *See* QUANTUM THEORY NONRELATIVISTIC

Quarrying

The process of extracting stone for commercial use from natural rock ledges The industry has two major branches A dimension stone branch involves the preparation of blocks of various sizes and shapes to be used as building stone for memorials and in various other ways The second major branch of the quarrying industry is the preparation of crushed and broken stone Different methods and equipment are used in the two branches

Dimension stone Although numerous and widely scattered only a small proportion of rock deposits satisfy the requirements for building and



Marble quarry with a block ready for hoisting and channeling machines in the background (*Vermont Marble Co*)

The transition from sandstone to quartzite is gra dational All stages of relic clastic structures are encountered Some sandstones are soon completely metamorphosed Others are very resistant and in many highly metamorphic quartzites of the Pre cambrian, there are relic structures still to be ob- served

Pure sandstones yield pure quartzites Impure sandstones yield a variety of quartzite types The cement of the original sandstone is in quartzite re- crystallized into characteristic silicate minerals whose composition often reflects the mode of devel opment Even the Precambrian quartzites corre spond to types that are parallel to present day de- posits

Carbonate cement reacts with silica to produce silicates during high temperature metamorphism However ferric silicate has never been observed The Fe_2O_3 pigment in deposits of desert sand re- . . . . . . h . Therefore old

Under the conditions of regional metamorphism cement composed of clay gives rise to sillimanite or kyanite potash rich cement yields potash feld spar or mica lime and alumina yield plagioclase or epidote dolomitic cement yields diopside or trem olite and siderite cement yields gruenerite Wol laston te w ll . . .

. . . e . . . g is marked however by thin strings of zircon iron ore or other inherited concentrations of the heavy min erals

Under the condition of contact metamorphism the cement of the original sandstone will recrys- tallize and minerals of the hornfels facies will de- velop Quartz itself is usually stable However in very hot contacts against basic intrusions quartz may invert to tridymite which again has reverted to

Waterloo quartz te showing schistos ty developed by shearing on the bedd ng planes Dodge County Wis consin (USGS)

quartz, or it may even melt Such vitrified sand stones resulting from partial fusion of inclusions in volcanic rocks are called buchites They often con tain fritted feldspar fragments corroded grains of quartz, and a matrix of slightly colored glass cor responding to the fused part of the sandstone See HORNFELS, see also QUARTZ, SILICATE MINERALS [T F W.B]

Quaternary

The latest major division of Cenozoic time (Ceno- zoic Era) the name refers collectively to all the geologic deposits (Quaternary System) that over lie or are younger than the Tertiary deposits and also to the time (Quaternary Period) during which

PRE CAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	CARBONIFEROUS		PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY
						Mississippian	Pennsylvanian						

these deposits accumulated The name was pro- posed by J D . . .
stand
ary a
Althougn Primary and Secondary have since been generally aband ned T . . .

. . . a Recent Series See CENOZOIC PLEISTOCENE, RECENT [R F F]

Quaternary ammonium salts

Analogs of ammonium salts in which organic radi cals have been substituted for all four hydrogens of the original ammonium cation Substituents may be alkyl aryl or aralkyl or the nitrogen may be part of a ring system

For . . .

. . . od of preparation The resulting quaternary ammonium salts are gen erally strong electrolytes and are converted to quaternary ammonium hydroxides very strong bases by silver oxide alcoholic potassium hydrox ide or anion exchange resins (see AMINE) Tri methylbenzylammonium hydroxide sold as Tri ton B catalyzes many condensation reactions of the aldol and Michael types.

A quaternary salt methylallylbenzylphenylam monium iodide containing four different groups on nitrogen proved to be resolvable (1899) clearly establishing that nitrogen is tetrahedral in these

Quartz

The most abundant and widespread of all minerals Quartz is an important rock forming mineral that occurs as a subordinate constituent of many igneous metamorphic, and sedimentary rocks, it is the principal constituent of sandstone and quartzite and of unconsolidated sands and gravels Quartz has the chemical composition SiO_2, silicon dioxide It crystallizes in the trigonal trapezohedral class of the rhombohedral subsystem The lattice type is hexagonal Quartz often occurs as well formed crystals The hardness is arbitrarily designated as 7 on the Mohs scale of relative hardness The specific gravity is 2 650 The fracture of crystals is subconchoidal, but more or less distinct cleavages are sometimes observed on the rhombohedrons $(10\bar{1}1)$ and $(01\bar{1}1)$ and on the prism $(10\bar{1}0)$ Ordinarily quartz is colorless and transparent, and the luster is vitreous Amethyst is a purple or bluish violet gem variety, and citrine is an orange brown gem variety, most citrine sold commercially is produced by heat treatment of amethyst Rose quartz is a massive type found in pegmatites Quartz also may have a smoky yellow to dark smoky brown color, varying to brownish black and almost opaque See AMETHYST

Properties Quartz is enantiomorphous, crystals are either right handed or left handed and correspondingly rotate the plane of polarization of transmitted light It also is piezoelectric Twinning is common in quartz. The chief twin laws are the Dauphine law in which the twinned parts are related by a rotation of 180° around the *c* axis and the Brazil law (or optical twinning) in which the twinned parts are enantiomorphs of different hand Quartz also known as α quartz or low quartz, is one of seven known polymorphs of silica It is stable below about 573°C, heated above this temperature it inverts reversibly to high quartz.

Large flawless crystals of quartz are employed for the manufacture of quartz oscillator plates, as prisms in optical spectrographs and as other optical devices Natural quartz crystals of commercial grade are principally obtained from Brazil but in recent years crystals of adequate quality have been synthesized by hydrothermal techniques See CRYSTAL OPTICS PIEZOELECTRIC CRYSTAL PIEZOELECTRICITY, POLARIZED LIGHT, SPECTROSCOPY

Varieties The varieties of quartz may be classed into two broad categories coarse crystalline including euhedral crystals and massive granular types in which the individual grains are visible to the unaided eye, and fine crystalline comprising massive types in which the individual grains or fibers are distinctly seen only under the microscope Coarsely crystallized quartz includes varieties based on color structure and inclusions The fine crystalline types of quartz usually have a waxy to dull luster, and are weakly translucent to opaque They are characterized by a slightly diminished specific gravity due to porosity and by greater ease of attack by chemical agents The fine-crystalline types of quartz are loosely divided into two main groups on the basis of particle shape These comprise the fibrous kinds, including chalcedony carnelian agate and their color variants and fine-granular kinds chiefly flint, together with jasper and its variants These materials have a very extensive varietal nomenclature that has arisen for the most part from their use since ancient times as gem or ornamental materials See AGATE, CHALCEDONY GEM, JASPER

[C.FR.]

Quartz clock

A clock that uses the piezoelectric property of a quartz crystal When a quartz crystal vibrates, a difference of electric potential is produced between two of its faces The crystal has a natural frequency of vibration depending on its size and shape, and if it is introduced into an oscillating electric circuit having nearly the same frequency as the crystal two effects take place simultaneously —the crystal is caused to vibrate at its natural frequency, and the frequency of the entire circuit becomes the same as the natural frequency of the crystal

In the clock the alternating current from the oscillating circuit is amplified and the frequency subdivided in steps such as from 100 kc to 1 kc finally driving a synchronous motor and gear train to display time by hands on a clock face

The natural frequency of a quartz crystal is nearly constant if precautions are taken in cutting and polishing it and if it is maintained at nearly constant temperature and pressure After a crystal has been placed in operation the frequency changes slowly, as a result of physical changes When allowance is made for such changes the best crystals run for a year with accumulated errors of less than 0 1 sec See OSCILLATOR

The practice at astronomical observatories that determine time is to rely on the average indication of several quartz clocks to carry the time from one astronomical determination to the next and to smooth out the accidental errors of the astronomical observations In this way the mean solar time is known at any instant with a precision of a few thousandths of a second See CLOCK

[C.M C.]

Quartzite

A metamorphic rock consisting largely or entirely of quartz. Most quartzites are formed by metamorphism of sandstone, but some have developed by metasomatic introduction of quartz, SiO_2 often accompanied by other chemical elements for example metals and sulfur (ore quartzites) The geological relations and the shape of quartzite bodies serve to distinguish between them The metasomatic quartzites are often found as contact products of intrusive bodies See METAMORPHIC ROCKS METASOMATISM, SANDSTONE

The transition from sandstone to quartzite is gradational. All stages of relic clastic structures are encountered. Some sandstones are soon completely metamorphosed. Others are very resistant and in many highly metamorphic quartzites of the Precambrian, there are relic structures still to be observed.

Pure sandstones yield pure quartzites. Impure sandstones yield a variety of quartzite types. The cement of the original sandstone is in quartzite recrystallized into characteristic silicate minerals whose composition often reflects the mode of development. Even the Precambrian quartzites correspond to types that are parallel to present day deposits.

Carbonate cement reacts with silica to produce silicates during high temperature metamorphism. However, ferric silicate has never been observed. The Fe_2O_3 pigment in deposits of desert sand resists any degree of metamorphism. Therefore old ... the same red color

... morphism

cement composed of clay gives rise to sillimanite or kyanite, potash rich cement yields potash feldspar or mica, lime and alumina yield plagioclase or epidote, dolomitic cement yields diopside or tremolite, and siderite cement yields gruenerite. Wollastonite will crystallize from pure calcite cement. Such quartzites occur as folded layers alternating with layers of ...

... olland illustrate this. True bedding is marked, however, by thin strings of zircon, iron ore or other inherited concentrations of the heavy minerals.

Under the condition of contact metamorphism the cement of the original sandstone will recrystallize and minerals of the hornfels facies will develop. Quartz itself is usually stable. However in very hot contacts against basic intrusions quartz may invert to tridymite, which again has reverted to ... or it may even melt. Such vitrified sand ... sions in ... n con ... ains of ... ss cor responding to the fused par ... ne. See HORNFELS; see also QUARTZ, SILICATE MINERALS.

(T F W B)



Waterloo quartzite showing schistosity developed by shearing on the bedding planes. Dodge County, Wisconsin (USGS)

Quaternary

The latest major division of Cenozoic time (Cenozoic Era), the name refers collectively to all the geologic deposits (Quaternary System) that overlie or are younger than the Tertiary deposits and also to the time (Quaternary Period) during which

PRECAMBRIAN		PALEOZOIC							MESOZOIC			CENOZOIC	
ARCHEOZOIC (EARLY PRECAMBRIAN)	PROTEROZOIC (LATE PRECAMBRIAN)	CAMBRIAN	ORDOVICIAN	SILURIAN	DEVONIAN	CARBONIFEROUS		PERMIAN	TRIASSIC	JURASSIC	CRETACEOUS	TERTIARY	QUATERNARY
						Mississippian	Pennsylvanian						

these deposits accumulated. The name was proposed by J. Desnoyers in 1829 as an addition to the standard names of rock groups (Primary, Secondary and Tertiary) that had been in use since 1760. Although Primary and Secondary have since been generally abandoned, Tertiary and Quaternary still remain in rather wide use. The Quaternary System is commonly subdivided into a Pleistocene Series and, above it, a Recent Series. See CENOZOIC, PLEISTOCENE, RECENT.

(R F F)

Quaternary ammonium salts

Analogs of ammonium salts in which organic radicals have been substituted for all four hydrogens of the original ammonium cation. Substituents may be alkyl, aryl or aralkyl, or the nitrogen may be part of a ring system.

For identification of tertiary amines and other purposes, direct alkylation with an iodide, sulfate or sulfonate (for example, methyl p-toluene sulfonate) is the preferred method of preparation. The resulting quaternary ammonium salts are generally strong electrolytes and are converted to quaternary ammonium hydroxides, very strong bases, by silver oxide, alcoholic potassium hydroxide, or anion exchange resins (see AMINE). Trimethylbenzylammonium hydroxide, sold as Triton B, catalyzes many condensation reactions of the aldol and Michael types.

A quaternary salt, methylallylbenzylphenylammonium iodide, containing four different groups on nitrogen, proved to be resolvable (1899), clearly establishing that nitrogen is tetrahedral in these

Quartz

The most abundant and widespread of all minerals Quartz is an important rock forming mineral that occurs as a subordinate constituent of many igneous metamorphic and sedimentary rocks, it is the principal constituent of sandstone and quartzite and of unconsolidated sands and gravels Quartz has the chemical composition SiO_2 silicon dioxide It crystallizes in the trigonal trapezohedral class of the rhombohedral subsystem The lattice type is hexagonal Quartz often occurs as well formed crystals The hardness is arbitrarily designated as 7 on the Mohs scale of relative hardness The specific gravity is 2 650 The fracture of crystals is subconchoidal but more or less distinct cleavages are sometimes observed on the rhombohedrons $(10\bar{1}1)$ and $(01\bar{1}1)$ and on the prism $(10\bar{1}0)$ Ordinarily quartz is colorless and transparent and the luster is vitreous Amethyst is a purple or bluish violet gem variety and citrine is an orange brown gem variety, most citrine sold commercially is produced by heat treatment of amethyst Rose quartz is a massive type found in pegmatites Quartz also may have a smoky yellow to dark smoky brown color varying to brownish black and almost opaque *See* AMETHYST

Properties Quartz is enantiomorphous, crystals are either right handed or left handed and correspondingly rotate the plane of polarization of transmitted light It also is piezoelectric Twinning is common in quartz The chief twin laws are the Dauphine law in which the twinned parts are related by a rotation of 180° around the *c* axis and the Brazil law (or optical twinning) in which the twinned parts are enantiomorphs of different hand Quartz also known as α quartz or low quartz is one of seven known polymorphs of silica It is stable below about 573°C, heated above this temperature it inverts reversibly to high quartz

Large flawless crystals of quartz are employed for the manufacture of quartz oscillator plates, as prisms in optical spectrographs and as other optical devices Natural quartz crystals of commercial grade are principally obtained from Brazil but in recent years crystals of adequate quality have been synthesized by hydrothermal techniques *See* CRYSTAL OPTICS PIEZOELECTRIC CRYSTAL PIEZOELECTRICITY POLARIZED LIGHT SPECTROSCOPY

Varieties The varieties of quartz may be classed into two broad categories coarse crystalline including euhedral crystals and massive granular types in which the individual grains are visible to the unaided eye, and fine-crystalline comprising massive types in which the individual grains or fibers are distinctly seen only under the microscope Coarsely crystallized quartz includes varieties based on color structure and inclusions The fine crystalline types of quartz usually have a waxy to dull luster and are weakly translucent to opaque They are characterized by a slightly diminished specific gravity due to porosity and by greater ease of attack by chemical agents The fine-crystalline types of quartz are loosely divided into two main groups on the basis of particle shape These comprise the fibrous kinds, including chalcedony carnelian agate and their color variants and fine granular kinds chiefly flint together with jasper and its variants These materials have a very extensive varietal nomenclature that has arisen for the most part from their use since ancient times as gem or ornamental materials *See* AGATE, CHALCEDONY GEM, JASPER

[C.F.R.]

Quartz clock

A clock that uses the piezoelectric property of a quartz crystal When a quartz crystal vibrates, a difference of electric potential is produced between two of its faces The crystal has a natural frequency of vibration depending on its size and shape and if it is introduced into an oscillating electric circuit having nearly the same frequency as the crystal two effects take place simultaneously —the crystal is caused to vibrate at its natural frequency, and the frequency of the entire circuit becomes the same as the natural frequency of the crystal

In the clock the alternating current from the oscillating circuit is amplified and the frequency subdivided in steps such as from 100 kc to 1 kc finally driving a synchronous motor and gear train to display time by hands on a clock face

The natural frequency of a quartz crystal is nearly constant if precautions are taken in cutting and polishing it and if it is maintained at nearly constant temperature and pressure After a crystal has been placed in operation the frequency changes slowly as a result of physical changes When allowance is made for such changes the best crystals run for a year with accumulated errors of less than 0 1 sec *See* OSCILLATOR

The practice at astronomical observatories that determine time is to rely on the average indication of several quartz clocks to carry the time from one astronomical determination to the next and to smooth out the accidental errors of the astronomical observations In this way the mean solar time is known at any instant with a precision of a few thousandths of a second *See* CLOCK

[G.M.C.]

Quartzite

A metamorphic rock consisting largely or entirely of quartz. Most quartzites are formed by metamorphism of sandstone but some have developed by metasomatic introduction of quartz SiO_2 often accompanied by other chemical elements for example metals and sulfur (ore quartzites) The geological relations and the shape of quartzite bodies serve to distinguish between them The metasomatic quartzites are often found as contact products of intrusive bodies *See* METAMORPHIC ROCKS METASOMATISM SANDSTONE

are zero then $q = s$ and is an ordinary *real number* or scalar Hence real quaternions include the real and ordinary complex numbers as special cases It will be evident that real quaternions are a kind of extension of the ordinary complex numbers $z = x + \sqrt{-1}\, y$

So far the case in which s a b c are complex [illegible] making q a

[illegible]

papers in which he assumed $i^2 = j^2 = k^2 = 1$ and thus paved the way to the dot or inner product of vector analysis Fundamentally quaternion algebra provides much deeper concepts and consequences and in some practical problems it presents clear advantages over vector analysis

Division In the division of quaternions reciprocals of the quaternion units i j k are easily found thus

$$i^{-1} = \frac{1}{i} = \frac{i}{ii} = -i \qquad j^{-1} = -j$$

$$k^{-1} = -k \qquad kk^{-1} = -k^2 = 1$$

More complicated quotients may be evaluated if care is taken to observe the defined conventions Thus

$$ij/j = ijj^{-1} = i$$

but

$$ji/j = jij^{-1} = -kj^{-1} = kj = -i$$

It is best to write denominators with negative exponents and place them properly in the numerator to avoid errors Similarly the reciprocals and quotients of real quaternions yield *unique results* If in $q = s + ia + jb + kc$ all of s a b c are not zero then

$$q^{-1} = 1/q = q' \quad qq' = q'/N, N = s^2 + a^2 + b^2 + c^2$$

Further if $p = u + ix + jy + kz$ is a second *real* quaternion then

$$p/q = pq^{-1} \quad qq = pq'/N$$

Accordingly real quaternions admit of division If $rq = p$ then $r = p/q = pq'/N$ and if $qr = p$ $r = q^{-1}p = q'p/N$ Hence both right division and left division yield unique quotients If for example $s = a = 1$ $b = \sqrt{-2}$ $c = 0$ then $N = 1 + 1 - 2 + 0 = 0$ this result is one reason why the discussion has been limited to real quaternions

Hamilton adopted the name vectors for directed lines in space If vectors are denoted with Greek letters, then $\alpha = ix + jy + kz$ is a vector whose components along a *conventional right handed* rectangular cartesian coordinate system are x y z respectively If $\beta = ix' + jy' + kz'$ is another vector in the same coordinate system then their products are

$$\alpha\beta = -(xx' + yy' + zz') + i(yz' - y'z) + j(x'z - xz') + k(xy' - x'y) = -u + \gamma$$

$$\beta\alpha = -(xx' + yy' + zz') - [i(yz' - y'z) + j(x'z - xz') + k(xy' - x'y)] = -u - \gamma$$

Both products yield a scalar u and a vector γ as a sum and such a sum is by definition a quaternion Further

$$\alpha\beta + \beta\alpha = -2(xx' + yy' + zz')$$

That is in general multiplication of nonparallel vectors is not commutative which is a special case of the noncommutation multiplication of quaternions If one sets $x = x'$ $y = y'$ $z = z'$ then

$$\alpha^2 = -(x^2 + y^2 + z^2)$$

which with negative sign is the square of the *length of the vector* α *whose components are* x y z

The quotient α/β emerges easily since

$$\alpha/\beta = -\alpha\beta/(x'^2 + y'^2 + z'^2)$$

Hence α/β has a unique value if x', y', z', are all real and not all zero The quotient of two vectors is a quaternion and previously it was stated that a quaternion can be defined as the ratio of two vectors

By multiplication a real q and its conjugate q' are found to commute

$$qq' = q'q = s^2 + a^2 + b^2 + c^2 = T^2$$

to give a real positive scalar Further, $q + q' = 2s$ and hence q and q' are the roots of a *quadratic* equation with real coefficients

$$t^2 - 2st + T^2 = 0$$

When this equation is solved in the field of ordinary complex numbers one finds

$$t = s \pm \sqrt{-(a^2 + b^2 + c^2)}$$

This simple but important result emphasizes the fact that in asking for a solution of a given algebraic equation the field of the quantities for which a solution is desired must be specified

Applications Next let $r = s' + ia' + jb' + kc'$ be another *real quaternion* then

$$\begin{aligned} qr = ss' &- (aa' + bb' + cc') \\ &+ i(as' + a's + bc' - b'c) \\ &+ j(bs' + b's + a'c - ac') \\ &+ k(cs' + c's + ab' - a'b) \\ &= S + iA + jB + kC \end{aligned}$$

where $S = ss' - (aa' + bb' + cc')$ $A = as' + a's - bc' - b'c$ etc for B and C Also

$$r'q' = S - (iA + jB + kC) = (qr)'$$
$$qr(qr)' = S^2 + A^2 + B^2 + C^2$$

But $qr(qr)' = qrr'q' = qq'rr'$ since rr' is a scalar and commutes with q and q' Therefore one has Euler's famous result that

compounds In general, optically active quaternary ammonium salts racemize readily, suggesting the possibility of an equilibrium between the tertiary amine and the alkyl compound from which they are generated

Surface activity of quaternary salts depends on the fact that many surfaces are negatively charged colloids Hence, positively charged cations are attracted to such surfaces Most commercial uses of these salts take advantage of this property Cationic surface-active agents of this type are rather obviously incompatible with anionic surface active agents such as soaps, since the high molecular weight cation and anion react to form a less soluble salt of still higher molecular weight Account must be taken of this fact in their manufacture and use as pharmaceuticals and disinfectants

Examples and uses of important salts are dimethylethyloleylammonium bromide, an antistatic agent for textiles (it may hold water on the surface) and a disinfectant in medicine and sanitation, 1 stearamidomethylpyridinium chloride, a water repellent for textiles, dimethyldioctadecyl ammonium chloride, a fungicide, emulsifier, paper softener, and corrosion inhibitor, tetraethylammonium chloride a ganglionic blocking agent, tubocurarine chloride a muscle relaxing agent, and choline bicarbonate, a parasympathomimetic agent

High molecular weight quaternary ammonium salts have proved useful in establishing theories of electrolytes particularly those involving a choice between dissociation and ion pair formation See AMMONIUM SALT, SURFACE ACTIVE AGENT [L B C]

Bibliography C A Lawrence *Surface Active Quaternary Ammonium Salts*, 1950

Quaternions

An associative noncommutative algebra based on four linearly independent units or basal elements Quaternions were originated in Dublin Ireland on October 16, 1843 by W R Hamilton (1805–1865), who is famous because of his canonical functions and equations of motion which are important in both classical and quantum dynamics

The four linearly independent units in quaternion algebra are commonly denoted by 1, i, j, k, where 1 commutes with i j k, and is called the principal unit or modulus These four units are assumed to have the following multiplication table

$$1^2 = 1 \qquad i^2 = j^2 = k^2 = ijk = -1$$
$$i(jk) = (ij)k = ijk$$
$$1i = i1 \qquad 1j = j1 \qquad 1k = k1$$

The i, j, k, do not commute with each other in multiplication that is $ij \neq ji$ $jk \neq kj$, $ik \neq ki$, etc But all real and complex numbers do commute with i, j, k, thus if c is a real number, then $ic = ci$, $jc = cj$, and $kc = ck$ On multiplying $ijk = -1$ on the left by i, so that $iijk = i(-1) = -i$, it is found, since $i^2 = -1$, that $jk = i$ Similarly $jjk = ji = -k$, when exhausted this process leads to all the simple noncommutative relations for i j k namely,

$$ij = -ji = k \qquad jk = -kj = i \qquad ki = -ik = j$$

More complicated products, for example $ikjk = -kki = i$, are evaluated by substituting for any adjoined pair the value given in the preceding series of relations and then proceeding similarly to any other adjoined pair in the new product, and so on until the product is reduced to ± 1, $\pm i$ $\pm j$ or $\pm k$ Multiplication on the right is also permissible, thus from $ij = k$, one has $ijj = kj$ or $-i = kj$ Products such as jj and jjj may be written j^2 and j^3

All the laws and operations of ordinary algebra are assumed to be valid in the definition of quaternion algebra, except the commutative law of multiplication for the units i, j, k Thus the associative and distributive laws of addition and multiplication apply without restriction throughout Addition is also commutative for example, $i + j = j + i$

Now if s, a, b, c, are real numbers rational or irrational, then a real quaternion q and its conjugate q' are defined by

$$q = s + ia + jb + kc \qquad q' = s - (ia + jb + kc)$$

In this case $qq' = q'q = s^2 + a^2 + b^2 + c^2 = N$ and N is called the norm of q, the real quantity $T = \sqrt{N}$ is called the tensor of q, and s, a b c are components (or coordinates) of q The part $\gamma = ia + jb = kc$ is the vector of q and it may be represented by a stroke or vector in a frame of cartesian coordinates a b c being its components Let now $p = u + ix + jy + kz$ be another real quaternion, if $pq = 0$ either p or q or both are zero, which is called the product law If for example, $p = 0$, then all of its components u, x, y, z are zero When $p = q$ that is $p - q = 0$ then one must have $w = s$, $x = a$, $y = b$ $z = c$, otherwise

$$(w - s) + i(x - a) + j(y - b) + k(z - c) = 0$$

would constitute a linear relation between 1, i j k which would be in conflict with their original definitions as linearly independent

Multiplication The product of two quaternions may be found by a straightforward process and in full is

$$pq = us - (ax + by + cz) + i(aw + sx + cy - bz) + j(bu + sy + az - cx) + k(cu + sz + bx - ay)$$
$$qp = us - (ax + by + cz) + i(au + sx - cy + bz) + j(bw + sy - az + cx) + k(cu + sz - bx + ay)$$

and hence

$$pq - qp = 2[i(cy - bz) + j(az - cx) + k(bx - ay)]$$

which is zero only when $cy - bz = az - cx = bx - ay = 0$ This shows that $pq \neq qp$ except under special conditions quaternion multiplication is not, in general, commutative

In q, if any two of a, b, c are zero one has in effect an ordinary complex number if all of a b c

can be represented on a plane The quantities $u = -\iota y$ $v = \iota\delta$ are also components of a unit spinor a function occurring in higher algebra and in the quantum theory (similarly for α β) If $a_x - (\frac{1}{2})(u^2 - \iota^2)$ $a_y = (-\iota/2)(u^2 + v^2)$ $a_z = uv$ are the components of a complex vector then $a_x^2 + a_y^2 + a_z^2 = 0$ and the vector is of length zero Such a vector is the starting point for some treatments of spinor theory

Returning now to the quaternion units 1 $\imath$ $\jmath$ k they may as appears to have been first discovered by Sylvester be represented by various matrices one 2×2 set being

$$1 \leftrightharpoons \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \imath \leftrightharpoons \begin{bmatrix} \sqrt{-1} & 0 \\ 0 & -\sqrt{-1} \end{bmatrix}$$

$$\jmath \leftrightharpoons \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \qquad k \leftrightharpoons \begin{bmatrix} 0 & \sqrt{-1} \\ \sqrt{-1} & 0 \end{bmatrix}$$

and the last three when multiplied by $\sqrt{-1}$ are the Pauli spin matrices occurring in the quantum theories of electron spin (spin $= \frac{1}{2}$) It has been shown by S Bochner that no set of 3×3 matrices have the multiplication table corresponding to 1 $\imath$ $\jmath$ k A set of 4×4 matrices due to A S Eddington which represents 1 $\imath$ $\jmath$ k is

$$1 \leftrightharpoons \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \imath \leftrightharpoons \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix}$$

$$\jmath \leftrightharpoons \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad k \leftrightharpoons \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}$$

These are four out of a group of sixteen 4×4 matrices used by Eddington in his fundamental theory five of the sixteen not including the above are when multiplied by $\sqrt{-1}$ the matrices occurring in Dirac s theory of the relativistic wave equation for the electron Another set of four matrices for 1 ι $\jmath$ k is

$$1 \leftrightharpoons \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad \iota \leftrightharpoons \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$\jmath \leftrightharpoons \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix} \qquad k \leftrightharpoons \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}$$

and if by ordinary matrix addition rules $x + \iota x_1 + \jmath x_2 + k x_3 = q$ are formed then

$$q = \begin{bmatrix} x & x_1 & x_2 & x_3 \\ -x_1 & x & -x_3 & x_2 \\ -x_2 & x_3 & x & -x_1 \\ -x_3 & -x_2 & x_1 & x \end{bmatrix}$$

and the determinant of q $|q|$ has the form

$$|q| = (x^2 + x_1^2 + x_2^2 + x_3^2)^2$$

an intriguing result

The basal quaternion elements 1 i $\jmath$ k and their negatives are the elements of a non Abelian group of order eight which is called the quaternion group The group contains proper self conjugate (invariant) cyclical subgroups one of the order 1 and significantly three of order 4 and there are five conjugate classes three of which are of order 2 and the others each of order 1

Hamilton's interest in analytical geometry led to many applications of quaternions in this field especially to conics and quadric surfaces In his hands and especially in those of P G Tait P Kelland J McAulay C J Joly and A S Hardy there have been many applications of quaternions to classical geometry and mathematical physics These include the use of Hamilton s partial differential operator

$$\nabla = \imath \frac{\partial}{\partial x} + \jmath \frac{\partial}{\partial y} + k \frac{\partial}{\partial}$$

The great advances made in quantum theory relativity number theory algebra and group theory are associated with scholars who had an easy familiarity with quaternions [D M Y]

Bibliography H F Baker *Abel s Theorem and the Allied Theory* 1897 H C Brinkman *Spinor Invariants* 1956 E J Cartan *Leçons sur la theorie des spineurs* 1932 C Chevalley *The Algebraic Theory of Spinors* 1954 L E Dickson *Algebras and Their Arithmetics* 1923 P Dienes *Taylor Series* 1931 A S Eddington *Fundamental Theory* 1946 H Goldstein *Classical Mechanics* 1950 H L Hamburger and M E Grimshaw *Linear Transformations* 1956 W R Hamilton *Lec*

heory of Numbers 3d ed 1955 H Jeffreys and B Jeffreys *Mathematical Physics* 3d ed 1956 C J Joly *Manual of Quaternions* 1905 P Kelland and P G Tait *Introduction to Quaternions* 3d ed 1904 L D Landau and E M Lifshitz *Quantum Mechanics* 1958 D E Littlewood *A University Algebra* 1950 C C MacDuffee *An Introduction to Abstract Algebra* 1940, F D Murnaghan in *A Collection of Papers in Memory of Sir William Rowan Hamilton* Scripta Mathematica 1945 A Sommerfeld Atombau und Spektrallinien Wellenmechanischer Erganzungsband 1929 P G Tait *Elementary Treatise on Quaternions* 3d ed 1890 E T Whittaker *Analytical Dynamics* 4th ed 1944

Quebracho

A name for a number of trees belonging to different genera but having similar qualities all indigenous to South America and valuable for both wood and bark The heartwood of one So h A—

known This is the world s most important source of tannin The logs are chipped and the chips are cooked with steam in

$$S^2 + A^2 + B^2 + C^2 = (s^2 + a^2 + b^2 + c^2)(s'^2 + a'^2 + b'^2 + c'^2)$$

which is important in number theory since it is used in the proof that every positive rational integer can be represented as a sum of four squares

The quotient of real vectors α/β is the sum of a scalar and vector and hence is a quaternion, for example, q If l_1, m_1, n_1, are the direction cosines of α in a rectangular cartesian coordinate frame, then

$$\alpha = a_0(il_1 + jm_1 + kn_1) \qquad a_0 = \sqrt{x^2 + y^2 + z^2}$$
$$\beta = b_0(il_2 + jm_2 + kn_2) \qquad b_0 = \sqrt{x'^2 + y'^2 + z'^2}$$

and

$$q = \frac{\alpha}{\beta} = \frac{a_0}{b_0}\,[l_1l_2 + m_1m_2 + n_1n_2 + i(m_2n_1 - m_1n_2) + j(l_1n_2 - l_2n_1) + k(l_2m_1 - l_1m_2)] = s + ia + jb + kc = s + \gamma$$

Then $\alpha = q\beta$, that is, q is an operator which turns and stretches β to coincide with α From analytical geometry $l_1l_2 + m_1m_2 + n_1n_2 = \cos\omega$, where ω is the angle ($<-$) between the vectors α and β It can be noted that

$$l_1(m_2n_1 - m_1n_2) + m_1(l_1n_2 - l_2n_1) + n_1(l_2m_1 - l_1m_2) = 0$$

and hence the vector part γ of q is perpendicular to α, and the same is true for γ and β This suggests the relationship

$$q = \frac{a_0}{b_0}(\cos\omega + \epsilon\sin\omega) = TUq$$

where $a_0^2/b_0^2 = s^2 + a^2 + b^2 + c^2 = T^2$, $\cos\omega = s/T$, and ϵ is the unit vector $\epsilon = (ia + jb + kc)/\sqrt{a^2 + b^2 + c^2}$, which is perpendicular to both α and β Because $\epsilon^2 = -1$, the square of any real unit vector is -1 The factor $Uq = \cos\omega + \epsilon\sin\omega$ turns β through the angle ω to coincide with the direction of α and is called the versor of q U is one of several symbols encountered in the grammar of earlier quaternion theory If $U'q = \cos\omega - \epsilon\sin\omega$ then $UqU'q = 1$

There is an important result in quaternion algebra which seems to be due to A Cayley With q as before then its reciprocal is $q^{-1} = q'/T^2$, as before If some other real quaternion is $p = w + ix + jy + kz$, then the expression qpq^{-1} is called the conical rotation (or transform) of p, and it finds important application in the specification of motions of rigid bodies We have

$$\begin{aligned} qpq^{-1} = \{&wT^2 + i[(s^2 + a^2 - b^2 - c^2)x + 2(ab - cs)y + 2(ac + bs)z] + j[2(ab + cs)x + (s^2 - a^2 + b^2 - c^2)y + 2(bc - as)z] \\ &+ k[2(ac - bs)x + 2(as + bc)y + (s^2 - a^2 - b^2 + c^2)z]\}/T^2 \\ = {} & w + iX + jY + kZ = w + \tau \end{aligned}$$

where $T^2 = s^2 + a^2 + b^2 + c^2$ If $T^2 = 1$, $u = w$ and

$$X = (s^2 + a^2 - b^2 - c^2)x + 2(ab - cs)y + 2(ac + bs)z$$
$$Y = 2(ab + cs)x + (s^2 - a^2 + b^2 - c^2)y + 2(bc - as)z$$
$$Z = 2(ac - bs)x + 2(as + bc)y + (s^2 - a^2 - b^2 + c^2)z$$

a linear transformation of X, Y, Z, to x y z

It is easy to show that, for example,

$$(s^2 + a^2 - b^2 - c^2)^2 + 4(ab - cs)^2 + 4(ac + bs)^2 = T^4 = 1$$
$$2(ab + cs)(s^2 + a^2 - b^2 - c^2) + 2(s^2 - a^2 + b^2 - c^2)(ab - cs) + 4(bc - as)(ac + bs) = 0$$

Hence the coefficients of x, y, z are direction cosines of the frame X, Y, Z, in the x, y, z frame of coordinates, and the transformation is orthogonal (unitary) The a, b, c, s are Euler's parameters (ξ, η, ζ, χ) Clearly the angle ψ between $\alpha = ix + jy + kz$ and $\gamma = ia + jb + kc$ is given by $g(ax + by + kz) = \cos\psi$, g a known quantity and some easy algebra shows that the angle between the vector part τ of qpq^{-1} and γ of $q = s + \gamma$ is given by the same expression Hence qpq^{-1} has rotated α conically about γ, and the magnitude of the rotation comes out, remarkably enough to be just 2ω, where again $\cos\omega = s/T = s$ As an example let

$$a = b = w = 0 \qquad s = \cos\omega \qquad c = \sin\omega$$

so $q = \cos\omega + k\sin\omega$, then the formulas above yield easily

$$X = x\cos 2\omega - y\sin 2\omega$$
$$Y = x\sin 2\omega + y\cos 2\omega$$
$$Z = z$$

which shows that p has been rotated conically about z and through the angle 2ω

If λ, μ, ν are the angles which the vector $\gamma = ia + jb + kc$ makes with x, y, z, respectively, and if $T^2 = 1$ then the relations $a = \cos\lambda\ \sin\omega$ $b = \cos\mu\ \sin\omega$, $c = \cos\nu\ \sin\omega$, $s = \cos\omega$, and $a^2 + b^2 + c^2 = \sin^2\omega$ are consistent with the above analysis Another formulation for s, a, b, c, is

$$a = \sin\frac{\theta}{2}\sin\frac{1}{2}(\psi - \phi) \qquad b = \sin\frac{\theta}{2}\cos\frac{1}{2}(\psi - \phi)$$
$$c = \cos\frac{\theta}{2}\sin\frac{1}{2}(\psi + \phi) \qquad s = \cos\frac{\theta}{2}\cos\frac{1}{2}(\psi + \phi)$$

The angles θ ϕ, ψ are the familiar Eulerian angles relating a fixed cartesian frame to another with the same origin the second being assumed in kinematics, to be fixed in a rigid body moving about a fixed point namely, the common origin of the two frames

If α, β, γ, δ are defined by $\alpha = s + ic$, $\beta = a + ib$, $\gamma = -a + ib$, $\delta = s - ic$, $i = \sqrt{-1}$, then $\alpha\delta - \beta\gamma = s^2 + a^2 + b^2 + c^2 = 1$,

$$\alpha = \cos\frac{\theta}{2}e^{i(\phi+\psi)/2} \qquad \beta = i\sin\frac{\theta}{2}e^{i(\phi-\psi)/2}$$
$$\gamma = i\sin\frac{\theta}{2}e^{i(\psi-\phi)/2} \qquad \delta = \cos\frac{\theta}{2}e^{i(-\psi-\phi)/2}$$
$$i = \sqrt{-1}$$

These are the Cayley Klein parameters which lend themselves to homographic (bilinear) transformations $z' = \alpha z + \beta/\gamma z + \delta$, in the complex z plane so that the motion of a solid body in space

time intervals are independent (3) the chance that a customer will arrive during any specified time interval of infinitesimal duration dt is $\lambda\, dt$ Expressed more exactly condition (3) means that the chance of one new arrival during any short time interval of length δt is $\lambda\,\delta t + o(\delta t)$, and the chance of more than one new arrival is $o(\delta t)$, as $\delta t \to 0$ Events (arrivals) having this character are said to constitute a Poisson process (stochastic process) Suppose that there is just one server and that if while he is occupied in serving a customer, further customers arrive, the latter will line up to await their turn for service in the order of their arrival Suppose that the time scale has been so chosen that the average duration of service of a customer is just 1 unit It will be assumed that customers do not all require the same time for service *but that their service times have what is known as an exponential distribution* such that the chance that a customer's service time will exceed any particular duration t is equal to e^{-t} This *service time distribution* has the remarkable property that whether or not a customer now being served will have his service completed during the next infinitesimal time interval of duration dt is *independent of how long he has already been served* the *chance of* completion is just dt It is assumed that *the service times* of successive customers are independent and that the number of *persons* waiting *has* no effect on the speed of service It *may be noted that* because the mean service time of a customer is equal *to 1 unit*, λ is equal to the traffic *intensity* measured in erlangs

The queue length L can be defined as the total number of customers at the counter so that $L = 0$ *if there* are no customers and someone is *waiting* for service *if* $L \geq 2$ and T denotes the length of time that a customer has to wait before his service begins *Let* $f_r(t)$ denote the *chance that* $L = r$ at time t During the infinitesimal time interval $(t, t + dt)$ there is the chance $\lambda\, dt$ that a new customer will arrive and, *provided that* $L \geq 1$ at time t there is the *chance* dt *that the customer at the*

... erential-difference equations

$$f_0'(t) = f_1(t) - \lambda f_0(t)$$

$$f_r'(t) = f_{r+1}(t) - (1 + \lambda) f_r(t) + \lambda f_{r-1}(t) \qquad (r \geq 1)$$

It can be proved that *if* $\lambda < 1$ the process approaches an *equilibrium or steady* state in which the chances $f_r(t)$ are independent of t and the derivatives on the *left hand* sides of the above equations vanish *The* equilibrium values are easily found to be

$$f_r = (1 - \lambda)\lambda^r \qquad (r \geq 0)$$

In particular *the chance* that at any arbitrary instant the server is idle is

$$f_0 = 1 - \lambda$$

For the mean queue length one finds

$$\mathcal{E}(L) = \sum_{r=0}^{\infty} r f_r = \frac{\lambda}{1 - \lambda}$$

Now consider a newcomer to the queue If there are L persons in the queue the instant before he arrives, the expected service time for each of these *(including the one* at the head) is 1, and so $\mathcal{E}(T) = \mathcal{E}(L)$ Thus the mean waiting time is

$$\mathcal{E}(T) = \frac{\lambda}{1 - \lambda}$$

This will be large if λ is close to, but less than 1 *If* $\lambda > 1$ it is obvious that the mean queue length increases indefinitely and *no equilibrium* is reached and it can be shown that the same *is true* if $\lambda = 1$

It has been assumed that the service times *follow* an exponential distribution *If instead it is assumed* that every customer requires *exactly the same time* for service namely 1 unit, a *somewhat similar* argument would lead *to the result that, for equilibrium* with $\lambda < 1$,

$$\mathcal{E}(T) = \frac{\lambda}{2(1 - \lambda)}$$

which is just one half the previous result *Both results are included* in a formula derived by F Pollaczek (1930) as follows If the service time distribution has mean equal to 1 and variance v and *if all the other* assumptions previously made *are satisfied then the* mean waiting time of a *customer, in* the equilibrium condition *for* $\lambda < 1$, is

$$\mathcal{E}(T) = \frac{(1 + v)\lambda}{2(1 - \lambda)}$$

This will tend to be large if either λ is close to 1 or v is *large (service* times are very variable)

Applications *Many* variants *of the* above problem have been stud...

a Poisson

patients a

spacing than a *Poisson process* if they have been given equally spaced *appointments* but their arrival times will *not be exactly* equally spaced The rule of queue *discipline* "first come first served" will not always apply Some customers, for example may enjoy priority *and move* ahead of non priority *customers There may* be more than one server and *if so separate* queues may be formed for each *There are then* various possible rules *to* be considered *for assigning* a newcomer to a queue It may happen that *the frequency* of arrivals or speed of service *varies* with the *queue* length, *for* example if the *queue* is too long new *customers* may be turned *away* Compound queueing systems are formed when *successive* services *have* to be rendered to the same customers

What is most interesting *to investigate* varies *with the* circumstances Sometimes it is the mean



Quebracho (*From P DeJanville, Atlas de Poche des Plantes Utiles des Pays Chauds, Librairie des Sciences Naturelles, 1902*)

copper extractors The liquor becomes highly concentrated with 40–60% of tannin It is used either alone or in combination with other chemicals for tanning all kinds of leather *See* SAPINDALES

[P D S]

Quellung reaction

A swelling of the capsule of a bacterial cell as a result of contact with an antiserum a serum containing antibodies, which are capable of reacting with the capsular polysaccharide The increase in volume is greater than could be accounted for on the basis of the amount of antibody which could be attached to the surface of the bacterial cell and is thought to be due to adsorption of water This reaction is sometimes employed in clinical practice as a method of identifying bacteria in sputum or spinal fluid especially the pneumococcus, Friedlander bacillus, and *Hemophilus influenzae* *See* BACTERIOLOGY, MEDICAL, HEMOPHILIC BACTERIA, IMMUNOLOGY, KLEBSIELLA PNEUMONIAE, PNEUMOCOCCUS

[P B BE]

Queueing theory

The mathematical theory of the formation and behavior of queues or waiting lines The name is also applied loosely to the mathematical study of a wide variety of problems connected with traffic congestion and storage systems Uneven flow through a service point with fluctuating arrivals and service times constitutes a major topic of operations research and some current work in queueing theory goes under that title For the mathematician queueing theory is particularly interesting because it is concerned with relatively simple stochastic processes which are in general non Markovian and possibly stationary *See* OPERATIONS RESEARCH STOCHASTIC PROCESS

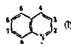
Origin. The principal pioneer of queueing theory was A K Erlang, who began in 1908 to study problems of telephone congestion for the Copenhagen Telephone Company He was concerned with problems such as the following A manually operated telephone exchange has a limited number (one or more) of operators When a subscriber attempts to make a call, he must wait if all the operators are already busy making connections for other subscribers It is of interest to study the waiting times of subscribers, for example, the average waiting time and the chance that a subscriber will obtain service immediately without waiting and to examine how much the waiting times will be affected if the number of operators is altered or conditions are changed in any other way If there are more operators or if service can be speeded up subscribers will be pleased because waiting will be reduced but the improved facility will be more expensive to maintain, therefore, a reasonable balance must be struck

Related problems arising today in the use of automatic telephone exchanges and of long distance lines able to carry only a limited number of messages simultaneously have resulted in much mathematical study of telephone traffic problems since Erlang's time His name is commemorated by telephone engineers in the unit of traffic intensity of a channel the erlang which may be defined as the number of requests for service (during some period of time) that are actually made divided by the number of requests that could have been satisfied if the channel had been used to full capacity the whole time

Similar problems arise in other contexts In a factory a number of machines such as looms may be under the care of one or more repairmen If a machine breaks down it must stand idle until a repairman is free from repairing other machines Machines here correspond to telephone subscribers breakdown corresponds to attempts to make a call and repair corresponds to connection Other examples of congestion situations mathematically similar or identical to the foregoing are aircraft flying around in circles waiting to use an airport landing strip automobiles lining up at a turnpike toll booth and (perhaps the most familiar of all) customers lining up at the counter of a retail shop waiting for service Much of the literature of queueing theory uses terminology appropriate to the latter example

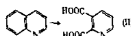
Theory One of the simplest portions of queueing theory will now be presented Consider customers arriving at a shop counter Suppose that they arrive singly in a purely random or haphazard fashion specified by the following conditions (1) the average number of customers arriving per unit time is constant, signified by λ, (2) the numbers of customers arriving during any two nonoverlapping

ing *quinolines* reference to the pyridine portion is made by the designation Py, and to the benzene portion by Bz Positions α β and γ are the same respectively as positions 2 3 and 4 Quinoline and some of its homologs are obtained as coal tar ex tractives The quinoline ring system appears in synthetic chemotherapeutic agents and in dyes Quinine and other natural alkaloids also contain the quinoline ring Quinoline itself is useful as a solvent as a source of nicotinic acid as an acid acceptor and dehydrohalogenating agent and as the starting point in organic syntheses

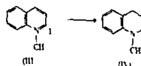


Properties Quinoline (I) is a colorless steam volatile liquid with bp 237 1°C mp −15°C spe cific gravity (15°) 1 09771 and n_D^{15} 1 62928 The solubility in water at 20°C is 0 65% Quinoline is a weakly basic (pK_a 4.85 at 20°C) tertiary amine that forms simple salts with acids and quaternary salts with alkylating agents The molecule is aro matic in chemical character It is resistant to dis ruption by heat alkali or acid it undergoes sub stitution reactions and it shows a resonance energy of 69 kcal/mole

Quinoline is oxidized by nitric acid and other reagents to quinolinic acid (II) In contrast, quinolinium quaternary salts are oxidized by alka line permanganate to give products of Py ring dis

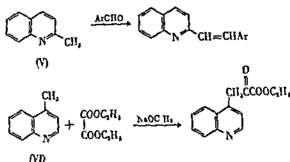


ruption Quinoline by either chemical or catalytic reduction gives 1 2 3 4-tetrahydroquinoline For the most part reactions of such tetrahydroquino lines may be interpreted as reactions of *N* alkyl orthosubstituted anilines *N* Methyl 1 2 3 4-tetra hydroquinoline or kairolin (IV) is formed by re duction of quinolinium methiodide (III)

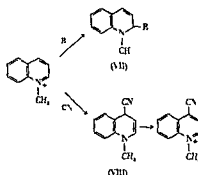


The quinoline 2 and 4 positions have properties that distinguish them from the other positions

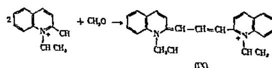
... ethyl oxalate to give a pyruvate derivative Halogens at positions 2 or 4 may be replaced more readily than those at other positions When the ni trogen carries a plus charge as in quinolinium quaternary salts these characteristic features be ... pronounced Quinolinium methiodide



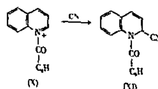
easily adds hydroxyl and alkoxyl radicals and car banions to give 1 2-disubstituted 1 2 dihydroquino lines (VII) Cyanide ion adds readily to the 4 posi tion to give the 1 4-dihydro derivative (VIII),



which can be oxidized to a 4-cyanoquinolinium salt 2 Methylquinolinium ethiodide condenses smoothly with aldehydes the product with formal dehyde represents one kind of cyanine dye (IX)



N Benzoylquinolinium cation (X) from the reac tion of quinoline and benzoyl chloride reacts eas ily with cyanide ion to give the useful Reissert compound (XI)



Although nitration conditions are known that will place the nitro group at every quinoline posi tion except 2 and 4 5-nitro- and 8-nitroquinoline predominate when nitric sulfuric acid is used Ni troquinolines serve as precursors to aminoquino lines Sulfonation at 100°C with fuming sulfuric acid gives quinoline-8-sulfonic acid as the main product

waiting time of customers sometimes the frequency with which the queue length exceeds a given limit sometimes the proportion of the servers time that is idle sometimes the average duration of a period during which a server is continuously occupied In the study of stocking a warehouse or retail shop known generally as the theory of inventories the frequency with which the stock will be exhausted is considered under various reordering policies Similar considerations apply in the theory of dams and water storage *See* LINEAR PROGRAMMING SYSTEMS ENGINEERING [F J A]

Bibliography A Doig A bibliography on the theory of queues *Biometrika* 44 (3 and 4) 490–514 1957 P M Morse *Queues Inventories and Maintenance* 1958

Quince

The quince (*Cydonia oblonga*) originated in Asia and is a deciduous crooked branched tree which grows to about 20 ft (*see* DECIDUOUS PLANTS) It is cultivated in either bush or tree form for its edible fruit The undersides of the leaves are densely tomentose (hairy) the solitary flowers up to 2 in across are snowy white or pale pink the fruit is a pear shaped or apple shaped pome characteristically tomentose up to 3 in in diameter aromatic sour astringent green turning clear yellow at maturity Used mostly for jam and jelly or as a stewed fruit it develops a pink color in cooking

The quince is propagated on hardwood cuttings or by budding on quince seedlings (*see* BUDDING) It may be used as stock for dwarfing the pear (*see* GRAFTING OF PLANTS PEACH PEAR) The tree is only slightly hardier than the peach the wood being severely injured at −15 to −20°F and is subject to fire blight (*Bacillus amylovorus*) which is controlled by removal of diseased twigs and avoidance of overly vigorous growth (*see* BACTERIA PLANT DISEASE) Trees are planted 15–20 ft apart Principal varieties are Orange Champion and Van Deman Fruits keep for only a few weeks There is little or no commercial production of the fruit in North America

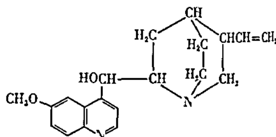
The dwarf Japanese quince *Chaenomeles japonica* with orange-scarlet flowers is grown as an ornamental shrub *See* FRUIT (TREE) FRUIT (TREE) DISEASES PLANT DISEASE CONTROL [H B T]



Quince (a) Flower borne on wood of the same season not from an autumn fruit bud (b) Fruit showing method of bearing at end of a contemporaneous shoot developed from a fruit bud of the previous season (From L H Bailey The Standard Cyclopedia of Horticulture vol 3 Macmillan 1937)

Quinine

The chief alkaloid of the bark of the cinchona tree which is indigenous to certain regions of South America The structure of quinine is



Its most important use has been in the treatment of malaria

For almost two centuries cinchona bark was employed in medicine as a powder or extract In 1820 P Pelletier and J Caventou isolated quinine and related alkaloids from cinchona bark and the use of the alkaloids as such rapidly gained favor Major credit is due to P Rabe for the postulation of the correct structure of quinine The difficult laboratory synthesis of quinine by R Woodward and W Doering in 1944 although economically unfeasible corroborated Rabe s structure

Until the 1920s quinine was the best chemotherapeutic agent for the treatment of malaria Clinical studies have definitely established the superiority of the newer synthetic antimalarials such as primaquine chloroquine and chloroguanide *See* ALKALOID CINCHONA MALARIA [S M K]

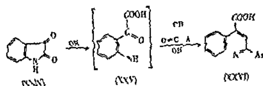
Quinoa

An annual herb *Chenopodium quinoa* 4–6 ft tall and a native of Peru is the staple food of many people in South America These plants grown at high altitudes produce large quantities of highly nutritious seeds used whole in soups or ground into flour which is made into bread or cakes They are also used as poultry food in medicine and in making beer The ash is mixed with coca leaves to flavor them as a masticatory In the United States the leaves are sometimes used as a substitute for spinach *See* CENTROSPERMALES SPICE AND FLAVORING [P D S]

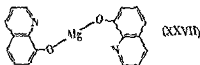
Quinoline

One of a group of organic compounds containing a benzene ring fused to the 2 3 positions of pyridine *See* HETEROCYCLIC COMPOUNDS PYRIDINE In nam

(XXVI) In the Pfitzinger synthesis one interpretation probably an oversimplification has the isatin molecule opening to form the o acylaniline (XXV) which then reacts with the acetophenone according to the Friedlander reaction

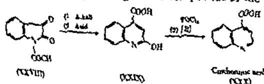


Important derivatives 8 Hydroxyquinoline (oxine) is prepared by sulfonation of quinoline followed by alkali fusion of the resulting quinoline 8-sulfonic acid 8 Hydroxyquinoline is a reagent of considerable utility in analysis for metals especially magnesium zinc and aluminum The procedures make ... of ... with met... oximate ... and an... line (vioform) and 7 iodo-8 hydroxy quinoline-5 sulfonic acid (chiniofon) are amebicides Chiniofon is also used in the colorimetric determination of iron and calcium

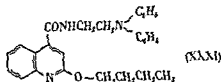


Quinoline carboxylic acids are prepared by several ring closure procedures by oxidation of groups such as methyl already on the ring or by transformations starting with bromoquinolines The monocarboxylic acids with pK_a 4 5–5 0 (in 50% methanol) are somewhat stronger than benzoic acid (pK_a 5 27) Quinoline 8 carboxylic acid (pK_a 7 2) is an exception The carboxyl groups especially at the 2 or 4 positions may be removed by heating

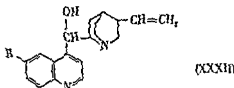
Quinoline 2-carboxylic acid (quinaldinic acid) is prepared by oxidation of 2 methylquinoline (quinaldine) or 2 styrylquinoline or by treatment of Reissert s compound (XI) with concentrated hydrochloric acid 4-Hydroxyquinaldinic acid (kynurenic acid) and 4 8-dihydroxyquinaldinic acid (xanthurenic acid) are two of the several metabolic products of tryptophan The cinchoninic acids (quinoline 4-carboxylic acids) are prepared by ring closures by oxidation of 4-substituted quinolines or by making use of compounds of the



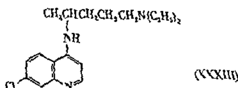
type (VIII) A convenient method for preparing cinchoninic acid is given in (XXVIII) to (XXX) Nupercaine (XXXI) a potent but somewhat toxic local anesthetic is prepared from cinchoninic acid derivative (XXIX) 2 Phenylcinchoninic acid (cincophen) prepared by Pfitzinger s isatin method has been used in treatment of gout and as an analgesic and antipyretic

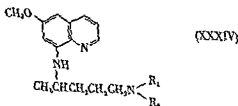


The cinchona alkaloids (XXXII) are quinoline derivatives in quinine and quinidine R = CH_3O in cinchonine R = H and in cupreine R = OH Quinine and to a lesser extent cinchonine have been used for centuries against malarial fever Quinidine a stereoisomer of quinine is a useful heart drug



Derivatives of 4-amino- and 8-aminoquinoline have attracted attention as synthetic antimalarial agents Chloroquine (XXXIII) one of the more effective 4-aminoquinolines is formed by combination of 4 7 dichloroquinoline with 4-amino-1 diethylaminopentane The 8 aminoquinoline derivatives (XXXIV) are prepared by alkylating 8-aminoquinoline with appropriate aminoalkyl halides



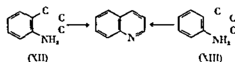


Pamaquine (XXXIV) in which $R_1 = R_2$ = ethyl isopentaquine in which R_1 = H and R_2 = isopropyl, and particularly primaquine in which $R_1 = R_2 = H$ are effective curative antimalarials The necessary 6-methoxy 8 aminoquinoline for (XXXIV) is obtained from 6-methoxy-8-nitroquin

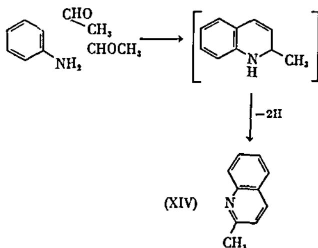
Bromination or chlorination of quinoline in the presence of sulfur occurs at the 3 position Bz haloquinolines are formed by ring syntheses starting with halogenated anilines Diazotization procedures can convert both Bz or Py aminoquinolines to the corresponding haloquinolines 2 Chloroquinoline is prepared by reaction of phosphorus chlorides with 2 quinolone or with 1 methyl 2 quinolone 4 Chloroquinoline is obtained by the action of phosphorus chlorides with 4-quinolone or by reaction of sulfuryl chloride with quinoline *N* oxide

Hydroxyquinolines other than 2 and 4-hydroxyquinolines resemble the naphthols in their general behavior, and may be regarded as normal phenols 2 Hydroxy and 4-hydroxyquinoline exist almost entirely in the quinolone form and show special behavior One property of considerable utility is the facile reaction of quinolones with phosphorus halides to give reactive 2 and 4-chloroquinolines 4-Quinolone or kynurine is the decarboxylation product from 4-hydroxyquinoline 2 carboxylic acid (kynurenic acid)

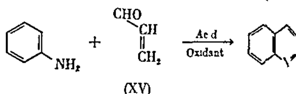
Preparation Quinoline ring syntheses construct the Py part of the quinoline product by operating on anilines or related compounds A useful classification of quinoline syntheses depends on whether the carbon atom of position 4 in the quinoline product was (XII) or was not (XIII) attached to the ortho position of the starting aniline



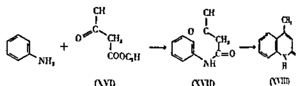
In the latter category (XIII) the Dobner Miller synthesis combines two molecules of aldehyde (reacting possibly as the derived bimolecular α β unsaturated aldehyde) with an aromatic amine In this way quinaldine or 2 methylquinoline (XIV)



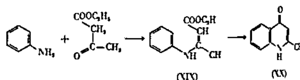
is obtained from acetaldehyde (that is presumably from crotonaldehyde) and aniline In common with some other quinoline syntheses the intermediate dihydro stage is not isolated but is dehydrogenated in situ In the versatile *Skraup synthesis* aniline and glycerol are heated in the presence of sulfuric acid and an oxidizing agent to form Py unsubstituted quinolines With the rationalization that glycerol acts as precursor to the α β unsaturated acrolein (XV) the Skraup reaction appears as a modification of the Dobner Miller process



Commercial synthetic quinoline is prepared by the Skraup reaction Syntheses with 1 3-dicarbonyl compounds are also possible In the Knorr quinoline synthesis aniline and acetoacetic ester (XVI)

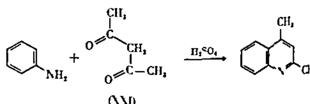


react to form acetoacetanilide (XVII) which with cold concentrated sulfuric acid cyclizes to give 4-methyl 2 quinolone (XVIII) The reactants can be made to combine in the opposite sense to give

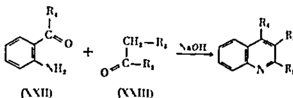


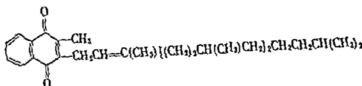
intermediate (XIX) which on pyrolysis furnishes 2 methyl 4-quinolone (XX) (Conrad Limpach method) When the 1 3 dicarbonyl compound is a diketone for example acetylacetone (XXI) the products are 2 4-disubstituted quinolines

Quinoline syntheses starting with compounds of type (XII) include the Friedlander method by

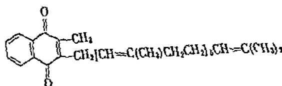


which an *o* acylaniline (XXII) condenses with an α methylene compound (XXIII) and the Pfitzinger scheme by which isatin (XXIV) combines with an acetophenone to give a 2 arylcinchoninic acid



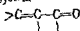


Vitamin K_1

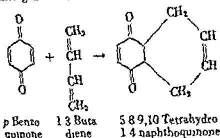


Vitamin K_2

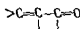
of the conjugated system



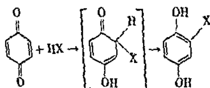
Reaction as a dienophile in the Diels Alder process is quite general and occurs under mild conditions



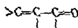
The remaining >C=C< bond of the quinone ring



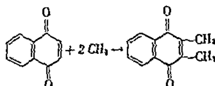
system by 1 4 addition and this is followed by enolization to a hydroquinone derivative Malonic ester



and acetoacetic ester react similarly with utilization of their active hydrogen atoms Mercaptans and Grignard reagents give mixtures of normal adducts to the C=O group and 1 4 additions to the



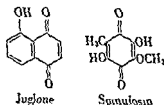
system Simple quinones do not often undergo substitution reactions with the electrophilic reagents commonly used for aromatic systems Free radicals from decomposition of acyl peroxides or lead tetra acetate substitute as follows



Methyl radicals

Important naturally occurring naphthoquinones are vitamins K_1 and K_2 which are found in blood and are responsible for proper blood clotting reaction The long aliphatic chain has been found unnecessary for the clotting reaction, its replacement by a hydrogen atom gives Menadione or 2 methyl 1,4-naphthoquinone which is manufactured synthetically for medicinal use

A number of quinone pigments have been isolated from plants and animals Illustrative of these are *juglone* found in unripe walnut shells and *spinulosin* from the mold *Penicillium spinulosum* 9 10-



Juglone Spinulosin

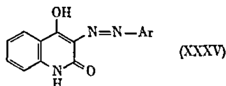
Anthraquinone derivatives form an important class of dyes of which alizarin is the parent type p Benzoquinone is manufactured for use as a photographic developer See ANTHRAQUINONE PIGMENTS, AROMATIC HYDROCARBON, HYDROQUINONE, KETONE

[D A S.]

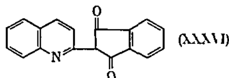
Bibliography L. F Fieser and M Fieser, *Organic Chemistry*, 3d ed , 1956

oline which is formed in a Skraup reaction with 2 nitro 4 methoxyaniline

The quinoline dyes include those of the cyanine type (IX) azo dyes (XXXV) derived from 2 4-dihydroxyquinoline and quinophthalones



(XXXV)



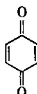
(XXXVI)

(XXXVI) from 2 methylquinolines and phthalic anhydride *See* ISOQUINOLINE [W J CE]

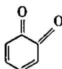
Bibliography R C Elderfield (ed) *Heterocyclic Compounds* vol 4 1952 R G W Hollingshead *Oxine and Its Derivatives* 4 vols 1954–1956 W E McEwen and R L Cobb The chemistry of Reissert compounds *Chem Revs* 55 511 519 1955 J P Phillips The reactions of 8 quinolinol *Chem Revs* 56 271 297 1956 E H Rodd *Heterocyclic Compounds* vol 4 1957

Quinone

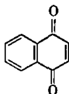
One of a class of aromatic diketones in which the carbon atoms of the carbonyl groups are part of the ring structure The name quinone is applied to the whole group but it is often used specifically to refer to *p* benzoquinone *o* Benzoquinone is also known but the meta isomer does not exist



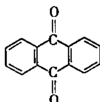
Quinone (*p* benzo quinone)



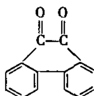
o Benzo quinone



1 4 Naphtho quinone



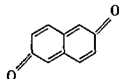
9 10 Anthraquinone



9 10 Phenanthra quinone

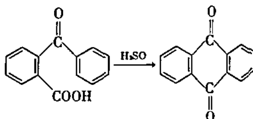
Preparation of quinones Quinones are prepared by oxidation of the corresponding aromatic ring systems containing amino (—NH_2) or hydroxyl (—OH) groups on one or both of the carbon atoms being converted to the carbonyl group *p* Benzoquinone is prepared by the oxidation of aniline with manganese dioxide MnO_2 in the presence of sulfuric acid H_2SO_4 Oxidation of phenol *p* aminophenol hydroquinone or *p* phenylenediamine will also produce *p* benzoquinone *o* Benzoquinone is prepared by oxidation of catechol with silver oxide Ag_2O in the absence of water This quinone is much less stable and more reactive than the para isomer

Three of the several possible quinones derived from naphthalene are known the 1 4 isomer shown above 1 2 naphthoquinone and 2 6-naphthoquinone.



2 6 Naphthoquinone

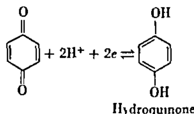
The naphthoquinones are prepared by oxidation of the corresponding aminonaphthols 9 10-Anthraquinone is best prepared by dehydration of *o* benzoyl benzoic acid which is prepared from Friedel Crafts reaction of benzene and phthalic anhydride Direct



9 10 Anthraquinone

oxidation of phenanthrene with chromic acid yields 9 10 phenanthraquinone the further oxidation of which gives diphenyl 2 2 dicarboxylic acid (diphenic acid)

Reactions of quinones *p* Benzoquinone is easily reduced to hydroquinone by a variety of reagents. The reaction is reversible and the position of equilibrium can be made to depend on hydrogen ion concentration and applied electrical potential



Hydroquinone

This system (E_0 = 0 699 volt) has been useful for the measurement of hydrogen ion concentration (*see* HYDROGEN ION) The E_0 values for many other quinone hydroquinone systems have been measured An intermediate in the reduction of *p* benzoquinone or in the oxidation of hydroquinone is quinhydrone a 1 1 molecular complex of these two substances

The most characteristic reactions of para quinones are those of the carbon to carbon double bonds and

R Rabbit to Rye

Rabbit

Any of many moderate sized soft furred short tailed jumping mammals of the family Leporidae order Lagomorpha present on all the major con



The cottontail rabbit *Sylvilagus floridanus* length to 15 in (From P M Duncan ed *Cassell's Natural History* Cassell)

tinents The family is roughly divided into the rabbits and hares The hares are born in a well furred advanced state of development with their eyes open rabbits are born naked with their eyes closed These are the most important of all North American game mammals in abundance number of individuals killed and the number of hunters seeking them See LAGOMORPHA [J D B]

Rabies

An acute encephalitic viral infection in man it is invariably fatal Human beings are infected from the bite of a rabid animal usually a dog See ANIMAL VIRUS

Infectious agent The virus is 100–150 millimicrons in diameter Canine rabies can infect all warm blooded animals and death usually results Some animals show chiefly paralytic signs whereas others manifest encephalitic hyperexcitability and [illegible] However the vampire bat may transmit [illegible] and [illegible] cur [illegible] grow [illegible] or in chick or mouse embryo tissue cultures Strains freshly isolated from dogs and man are called *street virus* whereas strains of altered pathogenicity and a stable shortened incubation period produced by serial passage in rabbit brains are called fixed Except in [illegible] street virus invariably produces cytoplasmic inclusion bodies (Negri bodies) in infected nerve cells See CULTURE EMBRYONATED EGG CULTURE TISSUE INCLUSION BODIES (VIRUS)

Pathogenesis The virus is believed to move from the saliva infected wound through sensory nerves to the central nervous system multiply there with destruction of brain cells and thus produce encephalitis with severe excitement throat spasm upon swallowing (hence hydrophobia or *fear of water*) convulsions and death—with paralysis sometimes intervening before death

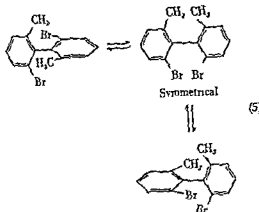
Diagnosis Diagnosis in the human is made by observation of Negri bodies in brains of animals inoculated with the patient s saliva or in the patient s brain after death Diagnosis in dogs is essential for guidance concerning human vaccination A dog which has bitten a person is isolated and watched for 7 days for signs of rabies if none occur rabies was absent If signs do appear the animal is killed and the brain examined for Negri bodies or for rabies antigen by testing with fluorescent antibodies

Epidemiology Rabies occurs throughout the world especially in India Africa and Europe Fewer than 100 human cases are reported annually in the United States but a reservoir is constantly present as shown by outbreaks in domestic and wild animals In South and Central America vampire bats and fruit and insect eating bats transmit the disease to cattle and occasionally to man

Prophylaxis The wound is cleansed immediately and sometimes cauterized with nitric acid The patient is vaccinated Indiscriminate human vaccination is inadvisable because there is risk of vaccination encephalitis Vaccination is avoided if the biting animal is found not to be rabid Such proof is available *with dogs* but inadequate in the case of bats for almost 50% of infected bats have no detectable Negri bodies Human vaccination is done with a series of doses of fixed virus The virus is chiefly the same strain that Louis Pasteur isolated in 1882 The hope is that antibodies will develop in the patient as a result of vaccination before the infecting street virus has had sufficient time to multiply Hyperimmune serum is also recommended it should be given as soon as possible after the bite See BIOLOGICALS

Control Rabies is controlled by compulsory vaccination of dogs and destruction of stray dogs cattle may also be vaccinated A living attenuated virus vaccine is now available

[J L M]



somers is called epimerization since epimers (diastereoisomers differing at one out of several asymmetric centers) result rather than a racemic form Any reaction resulting in racemization of a single-centered compound necessarily produces epimerization in a multiple centered compound See ASYMMETRIC SYNTHESIS, GLUCOSE, STEREOCHEMISTRY [W R V]

Rad

A unit of absorbed radiation dosage equal to the delivery of 100 ergs per gram of tissue The roentgen yields an ambiguous definition for the energy absorbed from radiation but rad makes this quantity definite See ROENTGEN UNIT [L Y CS]

Radar

An acronym for RAdio Detection And Ranging the original and still the principal application of radar The name is applied to both the technique and the ... —

c

r

... ELECTRICAL

Radar detection of an object is achieved by transmitting a beam of radio-frequency (rf) energy and detecting the energy reflected by the object Th...

by ...

th...

... near the transmitter

Radar ranging is accomplished by timing the period required for the rf energy to make the round trip Distance to the object is equal to one half the time elapsed times the velocity of the radio wave

Location of an object is obtained from the range and the azimuth and elevation angles of the radar antenna Many additional applications have been found and are mentioned later in this article

Radar has certain inherent advantages over detection systems employing light waves (1) it has greater range (2) it is usable in any weather and in day or night and (3) the electronic circuitry and components for transmitting receiving amplifying detecting and measuring are highly developed

Historical development. The fact that radio waves produce echoes was known before 1920 and the phenomenon was utilized to prove the existence of the ionosphere and to ascertain the height of its various layers by measuring the time required for a radio echo to return to the ground During the early 1930s various researchers in the United States England France, and Germany pointed out that ships and airplanes gave rise to radio echoes which could be used to deduce their location In 1935 the British Air Ministry authorized the installation of five radar stations on the east coast of England, and in 1937 fifteen stations more were added in order to provide radar surveillance of the air approaches to the entire east and southeast British coasts By the time World War II began an uninterrupted 24-hour radar watch was maintained over the principal air and sea approaches to Britain The radar network was so effective in locating German bombers and directing fighters against them that it is generally credited with making it possible for the severely outnumbered Royal Air Force to defeat the Luftwaffe in the Battle of Britain When the Germans resorted to night bombing in order to reduce the losses they suffered in daylight encounters, airborne radar aboard British fighter planes enabled them to train their guns on the ...

... orld

... oled

their knowledge to develop radar sets for such diverse applications as (1) search radar for surveillance of large regions and early warning of approaching ships aircraft and missiles, (2) fire control radar for tracking airborne or surface targets and automatically directing gunfire against them (3) aircraft intercept radar to be carried aboard airplanes for directing gunfire at enemy aircraft especially in the dark (4) radar bombsights and (5) airborne radar for submarine detection

Many nonmilitary applications have resulted from military developments Some of these are (1) marine and aircraft navigation radar, (2) airplane traffic-control radar for use in the control tower of busy airports (3) airport radar for directing airplane landings from the ground during conditions of extremely poor visibility, (4) radar altimeters (5) weather observation radar for discovering and tracking hurricanes, tornadoes and rainstorms (6) police radar for highway speed control and (7) tracking radar for monitoring the flight and obtaining geophysical data from space probes satellites and high altitude rockets See AIRBORNE RADAR, NAVIGATION SYSTEMS ELECTRONIC

Radar research and development since World War II has continued in order to meet the new requirements raised by long range rockets and space flight The principal areas of development are (1) radar for automatic guidance of unmanned vehicles (2) radar possessing very high precision in determining the position and velocity of targets at a range of thousands of miles and (3) reconnais-

Raccoon

An American carnivore, *Procyon lotor*, of the family Procyonidae, found from Panama to southern Canada, except for parts of the Great Basin



The raccoon *Procyon lotor*, length 30 in (From E L Palmer, *Fieldbook of Natural History*, McGraw Hill, 1949)

and the western mountains It is medium sized and mixed gray brown, and black in color The tail is ringed with black and there is a black mask across the face Its fur is of excellent quality and widely used, although not as much in demand now as in the 1920s

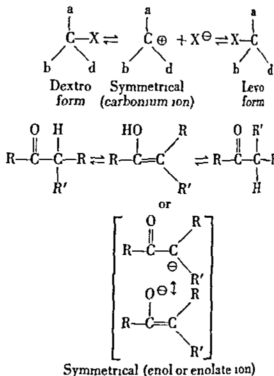
Raccoons are intelligent inquisitive animals They usually den in large hollow trees, but also will make their homes in holes dug in the soil They prefer to hunt along streams or lakes eating cray fish mollusks, and fish See CARNIVORA [J D B]

Racemization

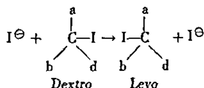
A racemic substance is composed of equal numbers of enantiomorphic dextro and levo forms and racemization is the transformation of either optically active enantiomorph into a racemic mixture or racemic compound (a 1 1 dextro levo molecular complex) See OPTICAL ACTIVITY

For racemization to occur, molecular dissymmetry must be lost, thus there must be possible an intermediate which is identical with its own mirror image Where a molecule possesses but one asymmetric carbon (C*), it must be possible to rehybridize the tetrahedral (sp^3) orbitals of C* to trigonal (sp^2) orbitals Thus the tetracovalent C* *must* become tricovalent The most common reactions permitting this are ionization to a carbonium ion Eq (1), and enolization, either to a true enol or to an enolate anion Eq (2) The planar carbonium ion of Eq (1) can then recapture an anion or solvent molecule from either side with equal probability to yield equal numbers of both dextro and levo forms Similarly in Eq (2), the enolic hydrogen (or proton lost by ionization) can return to either side of the original asymmetric carbon to produce a racemic modification

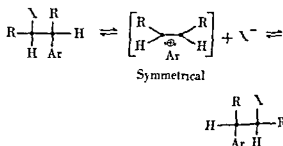
Related to the carbonium ion racemization but differing mechanistically is the Walden inversion in which a nucleophilic ion displaces an identical ion with simultaneous inversion of the molecular



configuration Racemization is complete when of the original molecules have reacted as show Eq (3)



Multiple centered (C*) dissymmetric c pounds may racemize via reactions which inv migration of a group provided that, in the co of migration, the molecule achieves molecular s metry (plane, center or 4-fold alternating axi symmetry) See Eq (4)



Molecular dissymmetry may also be lost w attendant racemization by the overcoming of energy barrier to free rotation upon which the dependent existence of enantiomorphs depen for example Eq (5) shows the racemization the optically active biphenyls

When a reversal of configuration occurs (as one of the above processes) at only one of seve asymmetric centers, the probability of exac equal quantities of each diastereoisomer bei formed is low since the intermediate is still d symmetric Such an interconversion of diastere

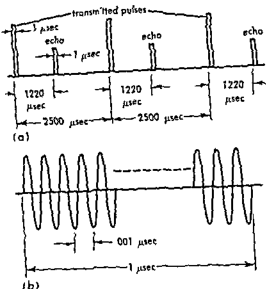


Fig 2 Waveforms in a typical pulse radar (a) Portion of transmitted pulse train and echoes Delay of 1220 μsec between start of transmitted pulse and echo corresponds to range of 100 nautical miles Interpulse period of 2500 μsec corresponds to pulse repetition frequency of 400 pps Echo amplitude is shown greatly exaggerated compared to amplitude of transmitted pulses (b) Single transmitted pulse showing carrier cycles. Carrier period of 0 001 μsec corresponds to L band frequency (1000 Mc)

output is continuous rather than pulsed no large power peaks occur

Radar carrier frequencies are broadly identified by a nomenclature which originated in wartime secrecy but which has persisted because of its convenience and general acceptance The spectrum is divided into bands the center frequencies and wavelengths of which are given in Table 1 See MICROWAVE

Radar propagation Radar propagation is governed by the laws of electromagnetic radiation A number of effects occur in radar because of the wa elen h —

because the ionosphere does not act as a reflector at the frequencies employed The *radar horizon* however is more distant than the optical horizon because the temperature and moisture gradient of the troposphere cause its dielectric constant to decrease with increasing altitude Radar

Table 1 Radar carrier frequency bands

Band	Center frequency	Wavelength
P	300 Mc	1 meter
L	900 Mc	33 cm
S	3 000 Mc	10 cm
C	5 000 Mc	6 cm
X	10 000 Mc	3 cm
K	"0 000 Mc	1 5 cm
Q	40 000 Mc	0 75 cm

rays are consequently refracted downward and the radar horizon is situated as though the earth were a sphere having a radius 4/3 times the actual value The limit of range for elevated targets imposed by ordinary radar propagation is through a fortuitous relation between units given by the formula

$$R_{max}\ (\text{statute miles}) = \sqrt{2h_1\,(feet)} + \sqrt{2h_2\,(feet)}$$

where R_{max} is the range limit in statute miles h_1 is the height of the radar antenna and h_2 is the height of the target in feet

Superrefraction Superrefraction sometimes causes the range limit to be remarkably extended The phenomenon which is also known as ducting occurs when a deviation from the usual temperature or humidity gradient of the troposphere produces an extended horizontal stratum in which the dielectric constant decreases with unusual sharpness with increasing altitude The radar ray is confined between the stratum and the earth s surface and travels beyond the ordinary radar horizon This effect usually occurs over the ocean where sharp moisture gradients are formed or over warm land that cools rapidly at dusk creating a sharp temperature gradient The phenomenon of superrefraction can also foreshorten the range limit if the dielectric gradient is such that the radar ray is refracted upward Superrefraction effects occur only with waves that are launched in a direction within a few degrees of the horizontal

Tropospheric attenuation Tropospheric attenuation can be attributed to two causes molecular absorption by resonant excitation of uncondensed gases and scattering by particles of dust and water drops in clouds fog and rain The principal molecular absorption effects are due to oxygen which exhibits a strong resonance at a wavelength near 0 5 cm and water vapor which exhibits a resonance at 1 35 cm These wavelengths are unsuitable for long range propagation but are useful where it is desired to confine the radiation to a limited locality The absorption due to water vapor

ca u co centration At wavelengths shorter than 3 cm this type of attenuation increases rapidly with diminishing wavelength until in the region below about 1 25 cm the attenuation even from moderate rain far exceeds that from uncondensed water vapor or oxygen However the attenuation of radiation passing through rain and fog at wavelengths longer than about 10 cm is negligible although enough scattering occurs to produce an echo

Ionospheric effects Radar propagation through the ionosphere depends upon the frequency employed and fluctuates in a somewhat unpredictable manner The ionosphere is charged with free electrons and ions. An electromagnetic wave traveling through a charged region undergoes absorption rotation of the direction of polarization and a change in the velocity of propagation The rotation

sance radar for mapping the earth's surface from high altitude aircraft and satellites Various forms of continuous wave radar as well as improvements in conventional pulse radar, are becoming available as a result of the vigorous efforts to meet new problems

For additional articles discussing radar, *see* CONTINUOUS WAVE RADAR, MONOPULSE RADAR

Fundamentals of operation Figure 1 shows a block diagram of a typical pulse radar set In most pulse radar systems a single antenna serves for both transmission and reception The transmitted pulse ends before the arrival of echoes The duplexer protects the sensitive receiver by disconnecting it from the antenna during the presence of the powerful transmitter pulse Upon the termination of the output pulse the duplexer disconnects the transmitter from the antenna and channels all the returning echo power into the receiver

The number of transmitter pulses per second is the pulse repetition frequency (prf) The duration of the output pulse is the pulse width The ratio of the pulse width to the period between pulses is the duty cycle of the radar, it usually is between 0 1% and 1% The time delay τ between transmitting a pulse and receiving an echo from an object at range R in miles is

$$\tau = \frac{2R}{c} \quad \text{seconds}$$

where c is the velocity of electromagnetic propagation (186 000 miles per second) The prf is established by the synchronizer which furnishes timing reference signals to the various display units and controls the operation of the modulator The modulator delivers power in the form of a pulse to the transmitter output tube which generates a signal at the radar carrier frequency for radiation by the antenna

The returning echo is fed to a mixer or first detector, for heterodyning with the output of the local oscillator (*see* HETERODYNE PRINCIPLE) This converts the carrier frequency to a much lower value, known as the intermediate frequency (i f) which is more suitable for amplification The i f is usually between 15 and 90 Mc 30 and 60 Mc being preferred values The frequency conversion does not affect the pulse shape The local oscillator employs a klystron or uhf triode depending on the frequency band in which the radar operates The mixer output is fed to the i f amplifier which increases the signal level by 70–120 db in conventional systems The envelope of the i f amplifier output is extracted by the second detector The resulting video pulse is fed to video amplifiers which increase the signal level for presentation on the output displays

The radar equation The strength of echo signals is related to the parameters of the radar system by the radar equation

$$W_R = W_T \frac{G^2\lambda^2\sigma}{(4\pi)^3R^4}$$

where W_R = received echo power
W_T = transmitted power
G = antenna gain
λ = radar carrier wavelength
σ = target s radar cross section
R = range (distance from radar to target)

A consistent set of units such as the mks system must be employed Typical values of transmitted power in pulse radar lie between 10 kilowatts and 10 megawatts However because the duty cycle is small the average power is typically between 0 1% and 1% of the peak power In continuous wave radar the average power is about the same as in pulse radar but since the transmitter

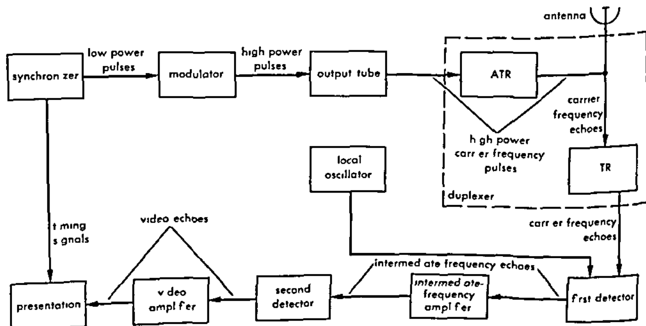


Fig 1 Block diagram of a conventional pulse radar This arrangement of components is used with either a search radar or a tracking radar with the appropriate type of antenna and presentation

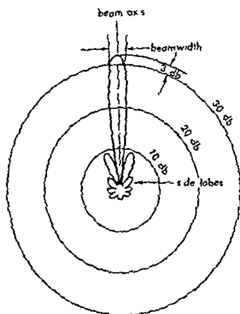


Fig 4 Typical search radar antenna beam pattern Antenna gain is plotted radially in decibels (db) versus angular distance from the beam axis The two points labeled A are the half power (3 db) points of the antenna radiation pattern

gain is one half the maximum value found on the axis With an illumination efficiency of 60% the antenna beamwidth θ is given approximately by

$$\theta \approx 70\frac{\lambda}{D} \quad \text{degrees}$$

where λ is the radar wavelength and D is the antenna diameter If the antenna aperture is not circular the beam will not have a circular cross section A search radar for example has a much longer aperture in the horizontal than in the vertical direction The beam formed by such an antenna will be narrow in azimuth and broad in elevation To calculate the horizontal or azimuth beamwidth the value of D in the above formula is the horizontal length while the vertical or elevation beamwidth is calculated by using the vertical length A typical search radar beam pattern is shown in Fig 4

Reflectors A large variety of techniques have been developed for generating the aperture illumination patterns required for various types of antennas One of the simplest and most frequently employed is the use of a parabolic reflector with the radiation source or feed placed at its focal point The radiation is reflected from the various points on the parabolic surface in a manner that forms a plane wavefront The operation is the same as that of a searchlight in which the light source is located at the focal point of a parabolic reflector and the reflected light emerges as a pencil beam of parallel rays This type of antenna of course cannot be used where the dimensions of the feed relative to the size of the reflector are so large that ... transmission line connected to it form an objectionable shadow in front of the reflector and perturb the radiation pattern However if the antenna diameter is more than fifteen times the wavelength the feed shadow is quite tolerable The parabolical reflector type of antenna has the advantages of simplicity and steerability and permits the polarization of the radar wave to be easily determined by controlling the shape and orientation of the feed

Dipole arrays Another frequently employed type of radar antenna consists of an array of dipoles arranged in a planar matrix to form the antenna ... [illegible]

dividual dipole radiation pattern is ... tropic but the relative amplitudes and phases of ... [illegible]

this spacing is close enough so that no discernible defect results from the fact that the aperture consists of an array of discrete radiators rather than a continuously illuminated surface The appearance of the assemblage is such that it resembles a bedspring thereby suggesting the name bedspring antenna by which it is commonly known This type of antenna was used in the earliest microwave radar in 1941 and is still frequently employed in the most recent designs It can be made fairly rugged and is easily steered in one dimension usually for horizontal scanning It is not usually rotated in the vertical direction because of the difficulty in achieving sufficient rigidity to maintain the relative orientations of the dipoles when the array is tilted This type of antenna is excited from the back therefore the problems of the feed shadow, which occur with reflector antennas, are avoided

Lens antennas Lens antennas of two types have been developed to provide easy steerability in both azimuth and elevation for tracking radars while avoiding the problems associated with the feed shadow in reflector antennas In a lens antenna the excitation is from the back and the lens serves to collimate (make parallel) and direct the radiation The velocity of propagation within the lens is different from that in air The lens is shaped so that the total time required for radiation to propagate from the feed point through the lens to a plane in front of the antenna is the same on all possible paths This condition assures the formation of a collimated plane wave

The two types of lenses employ dielectrics and metal plates, respectively The operation of dielectric radar lenses is the same as that of optical lenses The index of refraction within the dielectric is greater than in air and the lens is made thick at the center and thin near the edges The polarization of the radar wave is not affected by dielectric lenses Metal plate lenses are composed of parallel plates of metal like a venetian blind The region between adjacent plates is like a wave guide in that the phase velocity of an electromagnetic wave is determined by the relation between the

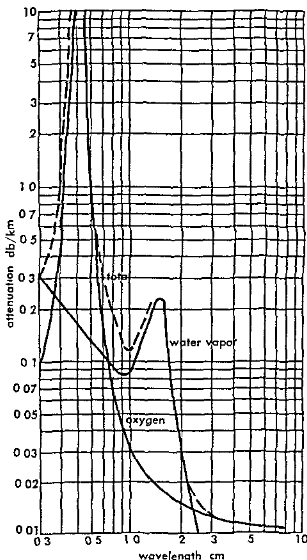


Fig 3 Atmospheric absorption versus wavelength The water vapor curve is for 10 grams per cub c meter (66% relative humid ty at 18°C) and the oxygen curve was taken on a sample of gas at 15 cm Hg pres sure

of the earth causes diurnal variations in the ioniza tion and height of the ionosphere above a given locality Fluctuations in the rate of ejection of par ticles from the sun also affect the state of the iono sphere Although the effect on radar waves of travel through a charged region can be calculated the unpredictability of the charge density and distribu tion makes the effect of the ionosphere difficult to [illegible] rve a three- [illegible] transmitted power in the direction of the target provide a large area to collect the echo power of the returning wave and indicate the angular position of the tar get

The beam pattern of a radar antenna is pro duced by diffraction from the antenna surface in a manner exactly similar to the generation of an optical diffraction pattern from an aperture illumi nated by light The beam pattern is formed in ac cordance with Huygens' principle The electro- magnetic field at each point on the antenna surface gives rise to a wavelet which travels away from the antenna at the speed of light The interaction of the wavelets generated at the various points on the antenna surface causes constructive interference in some directions and destructive interference in oth- ers In those directions in which the phases of the wavelets are mutually reinforcing a strong field is produced In the directions in which the pha es of the wavelets produce mutual cancellation the field is weak The radiation pattern of a radar antenna therefore consists of lobes of strong intensity sepa rated by nulls The same effects occur with light passing through an aperture However the lobes of a radar pattern are relatively broad because the ratio of the antenna diameter to the radar wavelength is relatively small hardly ever exceed ing 100 1 By analogy to optical diffraction phe nomena the antenna surface is referred to as an aperture and the manner of variation in amplitude and phase of the electromagnetic field across the surface is called the aperture illumination pattern

Gain The gain of an antenna varies with direc tion Gain is defined by reference to an isotropic radiator that is, a radiator that transmits power equally in all directions The gain of an antenna in a particular direction is the ratio of the power it sends in that direction to the power that would be sent by an isotropic radiator connected to an equally powerful transmitter The antenna gain for receiving is the same as for transmission When the antenna gain is referred to without specifying the direction the maximum value is usually im plied In antennas producing a symmetrical radia tion pattern the maximum gain usually occurs on the beam axis The maximum possible value of an tenna gain G_0 is given by

$$G_0 = \frac{4\pi A}{\lambda^2}$$

where A is the antenna surface or aperture area and λ is the radar wavelength This value of gain is actually attained by a uniformly illuminated aperture However most antenna apertures are not uniformly illuminated because it would lead to un desirably large minor lobes on both sides of the principal lobe in the beam pattern In order to re duce the side lobes the aperture illumination pat tern is tapered so that the electromagnetic field is stronger at the center of the antenna surface than at the edges In a well designed antenna the side- lobe power can be reduced by a factor of ten while at the same time the gain on the beam axis is de- creased 40–50% from the value of G_0 given above Side lobes must be reduced to suppress strong echoes from the ground and large targets other wise these echoes would interfere with weak echoes arriving through the principal lobe The ratio of beam axis gain to the value G_0 is the ill mination efficiency This is the ratio of power the antenna collects from an echo wave to that which would be collected by a uniformly illuminated aperture of the same size

Beamwidth The antenna beamwidth is defined as the angle between the two poi [illegible] of the princi pal lobe on either side of the

plitude and phase modulation of the echoes it is possible to infer the angular position of the target relative to the radar Range measurements inferred from the echo arrival time complete the data concerning the target s location in three dimensions

A conical scan radar pos esses such a narrow beamwidth that it might have difficulty in finding its target initially In ground or ship based operation either visual assistance or acquisition data from a nearby search radar can be provided However, when the conical scan system is carried aboard an airplane other acquisition data are *not* available and the time allowed for initial lock on is extremely limited because of the high velocities of aircraft A mode of operation known as spiral scan is then employed This entails varying the squint angle continuously from its normal value to a maximum value of possibly 60° In so doing the radar beam escribes a spiral and explores the large volume

echoes and directs the radar to return to the conventional conical scan operation and track the target

Automatic tracking A conical scan radar is usually combined with an automatic tracking arrangement The combination is able to follow a target completely automatically once it is properly locked on The essential elements of an automatic tracking system in addition to the radar itself are three feedback loops which constantly keep the radar

tion sensed by the radar and that corresponding to tracking loop output voltages The outputs of the feedback loops direct the radar where to look for the target The range error is obtained by measuring the difference in time of occurrence of the echo signal and a range gate obtained from the range-tracking loop The azimuth and elevation error signals are obtained from the amplitude modulation imposed on the echoes by the nutation if the nutation axis points precisely at the target there will be no amplitude modulation and no azimuth and elevation errors will be indicated.

Since an airplane target is constantly in motion it is necessary for the radar to follow it continuously in order not to fall too far behind for the tracking loops to accommodate the error Circuitry can be incorporated within the tracking loops to compute automatically the target s rate of change of position in the three dimensional coordinate system employed This enables the loops to predict the position of the target and is extremely useful for tracking the target and for directing gunfire or missiles to collide with the target

Track while scan systems These are employed

[illegible] of targets with only a [illegible] radar scan is not varied

as targets are acquired and tracked A separate tracking loop is assigned to each target and a large number of such loops receive their input signals from the same radar Each loop in the most common arrangement tracks its target in range and azimuth The range information is obtained from the time delay between the radar output pulse and the arrival of the echo The azimuth information is obtained from the antenna position at the time that the appropriate target provides an echo Each time the search radar scans past the target the appropriate tracking loop compares the values of range and azimuth it has already computed with those in

rate of change of these two coordinates The chief problem in track while scan systems is that data on each target are obtained only once during an antenna scan rather than continuously as in a conical scan system The slow data rate means that a rapidly maneuvering target can elude the tracking loop In addition when two targets cross or when the echo fades for a long period, precautions must be taken against confusion occurring within the tracking loops. The usual safeguard is to design the loops with long time constants so that they will continue to coast along the last clearly computed trajectory for a long time in the absence of unambiguous data However this makes it even easier for a highly maneuverable target to elude tracking. Despite these difficulties reasonable success has been attained in the development of track while scan systems and further work is being pursued applying the techniques of nonlinear time varying and adaptive loop design

Electronic scanning The scanning of a radar beam can be accomplished by purely electronic techniques which require no mechanical motion The chief advantage of electronic scanning is the much greater speed with which the [illegible]
changed fro
with Huyge[illegible]
cipal lobe is determined by the relative phases of the electromagnetic field across the antenna aperture In an antenna composed of an array of separate elements such as the bedspring antenna the excitation phase of each of the dipoles can be arranged to tilt the beam For example if the phase at the left side of the aperture is caused to lead the phase at the right the equiphase front of the radar wave is tilted rightward and the principal lobe is directed toward the right by an amount equal to the phase-front tilt In order to realize all the flexibility theoretically inherent in electronic scanning precise and rapid control of the excitation phase of each element of the antenna array is required A complex electronic system is necessary to achieve this control The advantages of rapid beam motion are especially important in using a single radar to perform track while scan with a large number of targets

Radar presentations A variety of displays has been developed for the presentation of radar output

plate spacing and the wavelength Since the phase velocity between the metal plates is faster than the velocity in free space the wave appears to travel faster while it is within the lens A metal plate lens therefore has a refractive index less than 1 in contrast to dielectric lenses in which the refractive index is greater than 1 A metal plate lens is shaped so that it is thin in the center and thick at the edges The polarization of a wave passing through a metal plate lens is usually made linear and parallel to the plates *See* ANTENNA (AERIAL)

Antenna scanning The geometric pattern described by the radar beam emerging from a moving antenna is called the scan pattern A search radar periodically scans a volume of the sky or a region on the earth s surface It is desirable to detect targets at the maximum possible range A search radar usually employs a beam that rotates in azimuth at a rate between 4 and 30 rpm The slower speed is likely to be used if weak echoes from distant objects are of special interest while the higher speed is likely to be employed if strong echoes from nearby rapidly moving targets are of prime interest An important case using slow scan speeds is that of searching for enemy aircraft detection at long range is important in order to provide early warning Rapid scanning rates are used for airport traffic control and marine navigation the expected echoes are likely to be quite strong and frequent observations must be made in order to provide instructions and avoid collisions

Fan beam In search radars a fan beam is employed This type of beam has a narrow (typically 1 5°) azimuth beamwidth and a broad (typically 45°) elevation bandwidth This beam shape is desired because azimuth resolution is usually much more important than elevation resolution because the azimuth position and range of an object divulge its geographical location which is generally of prime interest The elevation of targets is usually of lesser importance When target altitude is essential an additional or special antenna must be used The antenna used to form a fan beam is long in the horizontal dimension and relatively short in the vertical dimension

Cosecant squared pattern When a search radar is employed especially to maintain surveillance of aircraft a cosecant squared antenna pattern is desired This pattern distributes the radiated power in a manner that provides the greatest possible average detection range because it does not radiate any more than necessary in directions from which long range targets are not expected Because a definite altitude ceiling for airplanes exists it is not possible for long range targets to appear at high elevation angles Rather long range targets would be most likely to appear at low elevation angles A desirable radar antenna gain pattern is one in which all targets at the maximum altitude provide the same echo strength regardless of their angle of elevation as seen from the radar When the fourth power relationship in the radar equation is included it is found that the antenna gain should be proportional to the cosecant squared of the elevation angle The vertical gain pattern of a fan beam must be modified to conform with the cosecant squared relationship but the azimuth gain pattern should not be changed

Tracking radars such as those used for gunfire control missile guidance and measurement of satellite trajectories must produce three-dimensional position data The three coordinates employed are usually range azimuth angle and elevation angle Several scanning procedures have been developed to obtain three dimensional data.

Conical scan Conical scan employs a highly directive antenna which generates a narrow beam of circular cross section Either the feed or the reflector whirls rapidly in a manner that causes the beam axis to describe a circular cone this motion of the beam axis is called nutation The axis of the cone is known as the nutation axis and the angle between the antenna beam axis and the nutation axis is the squint angle (Fig 5) The squint angle is usually set equal to one half the antenna beamwidth but may exceed this figure Typical parameters of a conical scan radar are a beam width of 3° a squint angle of 1 5° and a nutation rate of 30 revolutions per second When the target lies directly on the nutation axis the antenna gain is the same in the target direction during all portions of the nutation cycle because the target always is displaced from the beam axis by exactly the same amount it is necessary for the beam to possess a perfectly symmetrical cross section for this to be true If the target does not lie precisely on the nutation axis the antenna gain in the target direction varies during the nutation cycle being largest when the beam axis comes closest to the target The echo amplitude then varies during the nutation c

echo is as a

frequency of

are dependent respectively on the angular displacement of the target from the nutation axis and the direction of the displacement By detecting the am

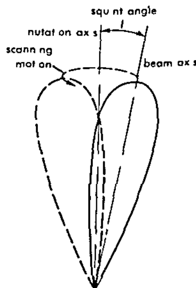


Fig 5 Schematic illustration of a conical scan antenna pattern and its motion

the circuit elements by a factor of thousands Examples of such circuits are multivibrators phantastrons and bootstraps See SAW TOOTH WAVE

The second general technique employs counting circuits in conjunction with a precise electronic clock The clock employs a stable oscillator containing a tuned circuit a quartz crystal or an atomic frequency standard from which a train of pulses is generated The occurrence of the transmitter output causes the pulse train to be applied to the counting circuit As soon as the preselected number of pulses has arrived the counting circuit produces a signal which indicates the elapse of the desired time delay relative to the transmitter output This signal can be used to start the sweep on the display See COUNTING CIRCUIT OSCILLATOR

Marker signals can be obtained from a saw tooth waveform generator or an electronic clock The saw tooth waveform is employed for this purpose in conjunction with a voltage discriminator circuit which indicates when the voltage applied to it reaches a preselected value The saw tooth voltage varies linearly with time therefore the indication by the voltage discriminator that the saw tooth has reached a certain value is tantamount to an indication that a known period of time has elapsed since the saw tooth began The discriminator indication can be applied as a marker causing a vertical deflection together with the echoes in which case this is known as an M display By using any of a number of easily produced waveforms the marker may be in the form of a step or a notch in the time base Alternatively the marker may be applied to the intensity control grid of the cathode ray tube so that it causes either a bright or dark spot

If the marker is obtained by use of a clock the output of the associated counting circuit can be applied to the display or the counting circuit can operate a gating circuit which passes the clock pulses directly to the cathode ray tube The gate [illegible]

[illegible] pulse train are presented

Type J display Another form of indication in which echoes prod ce defl[illegible]

J display [illegible]

oscillator [illegible]

chronized [illegible]

put of the [illegible] is applied to networks which produce two sinusoids in phase quadrature The two quadrature sinusoids applied to the vertical and horizontal deflection plates of a cathode-ray tube produce a circular pattern The angular position of the spot on the display at any instant is linearly related to the elapsed time since the instant of transmission A special cathode ray [illegible] is used [illegible] echoes a [illegible] pear as [illegible] frequency can be much higher than the prf of the radar in which case the intensity control grid of the cathode-ray tube is used to blank out the trace on all but one sweep cycle A delayed expanded presentation results

Plan position indicator (PPI) The output of a search radar can be presented in the form of a map showing the surface of the earth and airborne and surface targets on a plan position indicator, or PPI display The display presents echoes in polar coordinates A radial sweep starts at the center of the cathode ray tube when the transmitter pulse is emitted and moves toward the periphery of the tube The sweep arrives at the edge of the tube just as an echo arrives from a target at the maximum range of interest The angle of the radial sweep follows the beam axis of the scanning antenna The same type of linear sweep circuit as in the A display is used for the radial sweep The angular motion of the sweep is imparted by a servo mechanism connected to the antenna

Echo signals are applied to the intensity control grid of the cathode ray tube In the absence of an echo [illegible]

coordinate system can be supplied either by a transparent overlay or by adding electronic marker signals to the echoes applied to the intensity control grid The electronic range marks are obtained by applying a series of pulses with the appropriate time delays relative to the transmitter output They form a group of concentric circles at the ranges corresponding to their time delays The electronic angle marks are produced by brightening the spot for the entire duration of a few radial sweeps thereby producing a radial line The range marks are commonly placed so that three or four equally spaced circles are formed on the display and the angle marks are frequently located at every 10–30°

Two means of mechanizing the angular rotation of the cathode-ray tube sweep can be employed both of them obtaining the location of the antenna beam by a servomechanism attached to the antenna shaft The most frequently used method employs magnetic deflection of the beam to produce the radial spot motion The deflection is caused by the field of a coil or yoke, surrounding the neck of the cathode-ray tube The yoke in turn is rotated about the neck of the tube by the antenna servo This arrangement is simple and reliable Its only disadvantage is that the angular position of the spot cannot be changed discontinuously because [illegible]

[illegible] to the vertical and horizontal deflection plates are made proportional to the sine and cosine of the antenna beam angle This is accomplished by the use of a sine potentiometer or resolver synchro attached to the antenna servo In the dead time between echoes from targets at the maximum range and the emission of the next transmitter pulse other data can be presented on the display When this is done the

data (Fig 6) The choice for a particular application depends on the type of radar system and the information which must be inferred from it Almost all output displays employ cathode ray tubes The radar echoes produce either a deflection or an intensity modulation of the electron beam Sweep voltages synchronized with the antenna scan and the transmitter output position the electron beam so that one or more parameters such as range azimuth or elevation are mapped onto the display Cathode-ray tubes with a variety of different phosphor characteristics are available One important characteristic is that of persistence which is the time it takes the light output to decay to 10% of its initial value Phosphors having a short persistence are used for the display of rapidly changing or frequently repeated signals long persistence phosphors are employed when the period between signal repetitions is long or when it is desired to display a record of a target s trajectory as in the case of a search radar observing the course of an airplane The persistence of available phosphors may be from a few microseconds to more than a minute *See* CATHODE RAY TUBE

Type A display A frequently employed display for the presentation of radar outputs and of electronic waveforms in general is known as the A display The electron beam is swept horizontally across the face of the cathode ray tube at a uniform speed and the signal is applied to the tube so that it produces a vertical deflection The uniform sweep speed causes time to be mapped linearly onto the horizontal axis of the display The vertical deflection produced by the waveform results in a graph of the waveshape The horizontal sweep or time base can be started at a chosen instant thereby calibrating the time origin of the display The sweep rate can also be selected to produce any desired proportionality factor between time and the horizontal displacement of the spot Amplification of the signal before applying it to produce vertical deflection permits control of the proportionality factor between signal voltage and vertical displacement The A display known as an oscilloscope when it is not built into a radar presentation is one of the most widely used electronic instruments *See* OSCILLOSCOPE CATHODE RAY

Range is precisely measured by several forms of A display Accuracy is obtained either by the use of an accurately controlled time base or by the display of precise marker signals If the precision of the time base is relied upon means are provided to delay the start of the sweep by an accurately measured amount and to secure extreme linearity of the waveform used for horizontal deflection Delaying the sweep permits the signal to be examined in fine detail by spreading an echo of short duration across a large area of the display If marker signals are employed the time base need not be especially linear since the location of the marker signals and the echo are both affected identically In either case it is necessary to generate waveforms possessing characteristics which are precise consistent and linear functions of time

Two general means are available for the production of such waveforms One procedure generates a voltage or current varying linearly with time known as a saw tooth waveform Saw tooth waveform generators contain a circuit whose response to a step of voltage or current is an exponential waveform By making the exponential time constant very long the initial portion of the waveform closely approximates a linear function of time Feedback techniques are used to increase the time constant of

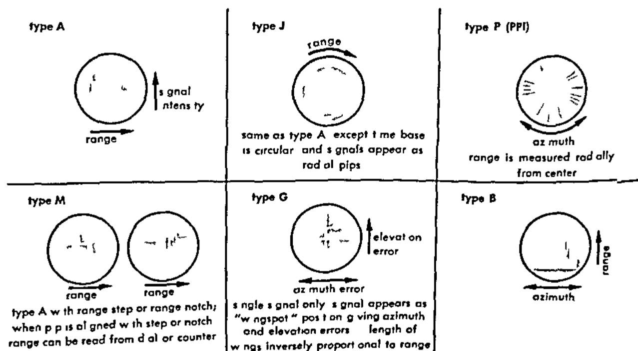


Fig 6 Radar displays

[illegible]e circuit elements by a factor of thousands [E]xamples of such circuits are multivibrators phan[t]astrons and bootstraps *See* SAW TOOTH WAVE

The second general technique employs counting [c]ircuits in conjunction with a precise *electronic [c]lock* The clock employs a stable oscillator con[t]aining a tuned circuit a quartz crystal or an [a]tomic frequency standard from which a train of [p]ulses is generated The occurrence of the *transmit[t]er* output causes the pulse train to be applied to the counting circuit As soon as the *preselected* number of pulses has arrived the *counting circuit* produces a signal *which indicates* the elapse of the desired time delay relative to the transmitter out*put This signal can be used to* start the sweep on *the display* *See* COUNTING CIRCUIT OSCILLATOR

Marker signals can be obtained from a saw tooth waveform generator or an electronic clock The saw tooth waveform is employed for this purpose in *conjunction* with a voltage discriminator circuit which indicates when the voltage applied to it reaches a preselected value The saw tooth voltage varies linearly with time therefore the indication by the voltage *discriminator* that the saw tooth has *reached a certain value* is tantamount to an indication that a known period of time has elapsed since *the* saw tooth began The discriminator indication can be applied as a marker causing a vertical deflection together with the *echoes* in which case this is known as an M display By using any of a number of easily *produced* waveforms the marker may *be in* the form of a step or a notch in the time base Alternatively the marker may be applied to the intensity control grid of the cathode-ray tube so that it causes either a bright or dark spot

If the marker is obtained by use of a clock the output of the associated counting circuit can be applied to the display or the counting circuit can operate a gating circuit which passes the clock pulses directly to the cathode-ray tube The gate serves the function *of passing* only the single clock pulse or small group of pulses of interest so that there *is no* question about which *members* of the clock pulse train are presented

Type I display Another form of indication in which [illegible]

[illegible] [illegible] [illegible] [illegible] instant of transmission The output of the *oscillator* is applied to networks which produce two sinusoids in phase quadrature The two quadrature sinusoids applied to the vertical and horizontal deflection plates of a cathode-ray tube produce a circular *pattern* The *angular* position of the spot *on* the display at any instant is linearly related to the elapsed *time* since the instant of transmission A special cathode-ray tube is used containing a central electrode to which the echoes are applied as negative signals Echoes appear as radial deflections The oscillator frequency can be much higher than the prf of the *radar in* which case the intensity control grid of the cathode ray tube is used to blank out the trace on all but one sweep cycle *A* delayed expanded presentation results.

Plan position indicator (PPI) The output of a search radar can be presented in the form of a map showing the surface of the earth and airborne and surface targets on a plan position indicator or PPI *display The* display presents echoes in polar coordinates A radial sweep starts at the center of the cathode ray tube when the transmitter pulse is emitted and moves toward the *periphery of the tube* The sweep arrives at the edge of the tube just as an echo arrives from a target at the maximum range of interest The angle of the radial sweep follows the beam axis of the scanning antenna The same type of linear sweep circuit as in the A display is used for the radial sweep The angular motion of the sweep is *imparted by a servo*mechanism *connected to the antenna*

Echo signals are applied to the intensity *control* grid of the cathode ray tube In the absence of an [illegible]

[illegible] coordinate system can be supplied either by a transparent *overlay or by adding electronic marker* signals to the echoes applied to the *intensity* control grid The electronic range *marks* are obtained by applying a series of pulses with the appropriate time delays *relative* to the transmitter *output* They form a group of concentric circles at the ranges corresponding to their *time* delays The electronic angle marks are produced by brightening the spot for the *entire* duration of a few radial sweeps thereby producing a radial line The range *marks* are commonly placed so that three or four equally spaced circles are formed on the display and the angle *marks* are frequently located at every 10–30°

Two *means* of mechanizing the angular rotation of the *cathode*-ray tube sweep can be employed both of them obtaining the location of the antenna beam by a servomechanism attached to the antenna shaft. The *most* frequently used method employs *magnetic* deflection of the beam to produce the *radial* spot motion The deflection is caused by the *field* of a coil or yoke surrounding the neck of the cathode ray tube The yoke in turn *is rotated* about the neck of the tube by *the* antenna servo This arrangement *is* simple and *reliable Its only dis*advantage is that the angular position of the spot cannot be changed *discontinuously because* of the inertia of the *yoke but this is not* ordinarily necessary The angular motion can also be obtained with *a cathode-ray tube* employing electrostatic deflection if *the sweep voltages* applied to the vertical and horizontal deflection plates are made propor*tional to the sine and* cosine of the antenna beam *angle* This is accomplished by the use of a sine potentiometer or resolver synchro attached to the *antenna servo In* the dead time between echoes *from targets* at the maximum range and the emission of the next transmitter pulse *other* data can be *presented on the* display When this is done the

data (Fig 6) The choice for a particular application depends on the type of radar system and the information which must be inferred from it Almost all output displays employ cathode ray tubes The radar echoes produce either a deflection or an intensity modulation of the electron beam Sweep voltages synchronized with the antenna scan and the transmitter output, position the electron beam so that one or more parameters such as range, azimuth or elevation are mapped onto the display Cathode ray tubes with a variety of different phosphor characteristics are available One important characteristic is that of persistence which is the time it takes the light output to decay to 10% of its initial value Phosphors having a short persistence are used for the display of rapidly changing or frequently repeated signals, long persistence phosphors are employed when the period between signal repetitions is long or when it is desired to display a record of a target's trajectory as in the case of a search radar observing the course of an airplane The persistence of available phosphors may be from a few microseconds to more than a minute *See* CATHODE RAY TUBE

Type A display A frequently employed display for the presentation of radar outputs and of electronic waveforms in general is known as the A display The electron beam is swept horizontally across the face of the cathode ray tube at a uniform speed and the signal is applied to the tube so that it produces a vertical deflection The uniform sweep speed causes time to be mapped linearly onto the horizontal axis of the display The vertical deflection produced by the waveform results in a graph of the waveshape The horizontal sweep or time base can be started at a chosen instant thereby calibrating the time origin of the display The sweep rate can also be selected to produce any desired proportionality factor between time and the horizontal displacement of the spot. Amplification of the signal before applying it to produce vertical deflection permits control of the proportionality factor between signal voltage and vertical displacement The A display known as an oscilloscope when it is not built into a radar presentation is one of the most widely used electronic instruments *See* OSCILLOSCOPE CATHODE RAY

Range is precisely measured by several forms of A display Accuracy is obtained either by the use of an accurately controlled time base or by the display of precise marker signals If the precision of the time base is relied upon means are provided to delay the start of the sweep by an accurately measured amount and to secure extreme linearity of the waveform used for horizontal deflection Delaying the sweep permits the signal to be examined in fine detail by spreading an echo of short duration across a large area of the display If marker signals are employed the time base need not be especially linear since the location of the marker signals and the echo are both affected identically In either case it is necessary to generate waveforms possessing characteristics which are precise consistent and linear functions of time

Two general means are available for the production of such waveforms One procedure generates a voltage or current varying linearly with time known as a saw tooth waveform Saw tooth waveform generators contain a circuit whose response to a step of voltage or current is an exponential waveform By making the exponential time constant very long the initial portion of the waveform closely approximates a linear function of time Feedback techniques are used to increase the time constant of

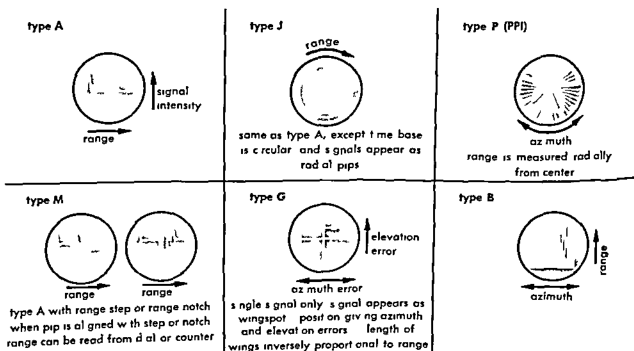


Fig 6 Radar displays

plitude and frequency of the local oscillator output and should not constitute a major contribution relative to the other noise sources in a well designed receiver

External noise The principal sources of external noise are solar and galactic radiation, thermally produced radiation from warm objects on the ground or in the sky and radiation from arcs sparks and corona in fluorescent lamps automobile

affect it because radiation at radar frequencies is not reflected by the ionosphere

Noise figure The noise figure is the figure of merit by which the noise quality of a receiver is measured A receiver composed of ideal components would suffer from no internal noise source except the thermal noise produced by the circuit connected to the antenna It is impossible to eliminate moleculer thermal agitation and the noise thereby produced except by cooling the circuit to absolute zero temperature However it is theoretically possible to eliminate all other sources of noise and an ideal receiver may be imagined in which this is done The *noise figure* of an actual receiver is the ratio of the noise power it produces to the noise power produced by an *ideal receiver* at room temperature possessing the same gain and frequency characteristics Practical values of noise figure depend upon the frequency in question because different circuit components are available at the various frequencies In the range above 1500 Mc a noise figure of 6–8 db is good and a value of 11 db is common Between 300 and 1500 Mc a noise figure of 4 or 5 db is good and 6 db is common Between 30 and 300 Mc a noise fig

[illegible]

MASER PARAMETRIC AMPLIFIER

Radar signal detection and accuracy The detectability of a signal and the accuracy of measurements made with it both depend on the signal-to-noise power ratio at the point in the radar system where the signal is displayed or used to reach a decision The instantaneous signal to-noise ratio at the receiver front end can be improved by various types of processing all of which require integration of successive samples of the signal in some manner Th

[illegible]

conf [illegible] the radar system or the type of processing employed the *theoretically best achievable* signal to-noise ratio is equal to E/N_0, where E is the total signal energy collected at the front end of the receiver during the period of integration and N_0 is the noise power per cps of bandwidth at the receiver front end due to both internal and external noise sources. In order to attain the theoretical optimum it is necessary to employ linear processing and mathematically ideal integration

In situations where the signal to-noise ratio is substantially better than unity before processing the ideal result can be approached very closely with nonlinear processing An illustrative case is that of video integration in which the signal is applied to a rectifier which passes only the amplitude modulation and rejects the sinusoidal carrier With pulse radar the amplitude modulation consists of pulses Successive pulses are added in an integration circuit or by superposition on the phosphor of a cathode ray tube The improvement in signal to noise ratio over that available with a single pulse is equal to the number of pulses integrated

If the signal to-noise ratio before processing is substantially less than unity it is necessary to employ coherent integration The signal must not be subjected to any nonlinear operation such as envelope or phase detection or passage through a nonlinear amplifier If a nonlinear operation occurs, it will cause heterodyning between the signal and noise resulting in a severe diminution of signal power The frequency carrying the signal can be shifted to a convenient value however by use of linear converters and stable local oscillators Coherent integration can be accomplished by applying the signal and noise to a *narrow bandwidth tuned* circuit or its equivalent, which resonates at the frequency of the carrier bearing the signal Successive portions of the signal reinforce coherently and continually add energy to the resonant circuit If the internal losses of the resonant circuit are sufficiently small the power ratio between the components due to signal and noise in the output will be equal to E/N_0 The term coherent integration derives from the utilization of signal phase continuity or *coherence* as compared to the random pha [illegible]

[illegible] ould produce the same intensity echo regardless of its orientation relative to the incident radar wave A close approximation to the ideal is obtained with a corner reflector which consists of three mutually perpendicular conducting plates forming a corner The incident wave bounces between plates and the *returned* echo varies only slightly as a function of the angle between the axis of the corner and the radar line-of sight Corner reflectors are used to test radar systems and to serve as conspicuous radar markers of geographic reference points and surveying bench marks

Most targets depart radically from the ideal The complex shape of objects such as aircraft causes the echo to fluctuate by many orders of magnitude as the aspect presented to the radar changes The effective reflectivity at any instant is described in terms of the radar cross *section* σ (see earlier section on the radar equation) defined as the area that would intercept that amount of radiation which, when reradiated isotropically produces an echo equal to that observed from the target The radar cross section fluctuates as the target aspect changes

ability of the electrostatic deflection plates to move the spot rapidly to any part of the tube is utilized

Search radars installed aboard ships or aircraft require a means of stabilizing the presentation so that turning motions of the vehicle do not result in a rotation of the display on the cathode ray tube Without stabilization, turning motions cause the bearings of all objects to change relative to the vehicle thereby smearing the presentation Stabilization is accomplished by the use of gyroscopes, which detect the vehicle's turning motion and apply a correction to the servomechanism controlling the angular position of the sweep The position of North on the display remains fixed and target echoes are presented at their true bearings rather than their relative bearings The display may also be stabilized with respect to the linear motion of the vehicle by sensing its displacement and correcting the location of the center of the presentation so that it remains fixed with respect to the earth

A particular sector of the region surveyed by the radar may be of special interest in which case it is desirable to expand the presentation of that area This may be done by offsetting the center of the display so that the position of the radar is at one edge The entire presentation then is devoted to the sector of interest, and the remaining area is not shown at all This is called offset center display If it is anticipated when the radar is designed and installed that only a particular sector will be of interest, the antenna scan can be arranged so that it scans only that sector instead of rotating 360° This sector scan can be accomplished by having the antenna oscillate back and forth over the same sector, but it is necessary to design the antenna mount to accommodate the angular accelerations that occur An alternative which permits uniform rotation of the mount is to place several apertures on it facing in different directions and connect the transmitter and receiver to each one as it bears upon the sector of interest

Delayed sweep is used when the region near the radar is of small interest or cannot be seen well because of nearby ground clutter, it is also commonly used in airborne search radar where the sweep is delayed until the first ground echoes are received Instead of starting the radial sweep on the PPI display when the transmitter pulse occurs the sweep is delayed An annular zone starting at the distance from the radar corresponding to the delay is presented and the radial scale is expanded because more space on the display is available for it Delayed sweep may be combined with an offset center or sector scan

Type B display The B display can be employed when delayed sweep is used together with an offset center or sector scan A B display is a presentation in rectangular coordinates of a zone containing an angle of 10–40 degrees and a range interval equal to 5–25% of the maximum range The display is obtained by plotting range and azimuth orthogonally Its advantage is that it permits close examination of a small region of particular interest

It is common to utilize the output of a large search radar to provide a number of different types of displays simultaneously This can be done without mutual interference

Three dimensional displays Three dimensional displays show range, azimuth, and elevation One of the most interesting, the G display, is for use in the cockpit of an airplane employing a conical scan radar The azimuth and elevation of the target relative to the airplane are indicated by the location of a spot in a display in which azimuth and elevation angles are plotted orthogonally and the range of the target is indicated by the length of two horizontal lines on both sides of the spot As the range decreases the lines are made longer, so that the presentation resembles a picture of an approaching airplane

Noise. The usability of a radar echo is determined by the signal to-noise ratio Noise is a random fluctuation of voltage or current The instantaneous values of noise are unpredictable, and so it is necessary to describe noise characteristics in statistical terms Radar noise can be classified according to its origin, either internal to the receiver or external Internal noise is generated by the receiving equipment, while external noise arrives at the antenna already mixed with the echo

Internal noise The principal sources of internal noise are thermal noise, vacuum tube noise, converter noise, and local oscillator noise

Thermal noise is caused by molecular thermal agitation within circuit components This source of noise produces the same amount of power per unit of bandwidth at all frequencies, noise possessing this characteristic of uniform spectral density is referred to as white noise The amount of thermal noise power delivered by a circuit depends upon its bandwidth and temperature *See* NOISE, ELECTRICAL

Vacuum tube noise is due to a number of causes The shot effect occurs because the electron stream is composed of discrete electrons whose individual arrivals at the anode result in minute fluctuations of output current Partition noise occurs in multielectrode tubes because of the discrete and random manner in which electrons are affected by various electrodes Both the shot effect and partition noise have the spectral characteristics of white noise The flicker effect arises because of slow variations in the emission characteristics of the cathode and the spectrum of this noise contribution is generally concentrated in the frequency region below 1000 cps *See* VACUUM TUBE

Converter noise occurs in the circuit of the receiver in which the radar carrier frequency is converted to a lower frequency in order to facilitate subsequent amplification and filtering Crystal converters contribute white noise which is probably [illegible] ulti [illegible] oise

Local oscillator noise occurs in the circuit that supplies the heterodyning frequency to the converter The noise is due to fluctuations in the am

ast 6–10 pulses should be transmitted during the me a *target is between* the half power points of he antenna pattern The same criterion applies f the essential *requirement* is to measure the an ular *position* of the target with maximum accu acy On the other hand only two or *three pulses per beamwidth* are required if the main purpose is *to perform* track while-scan on rapidly moving tar gets at fairly short range

The radar receiver is optimally adjusted *for ex* cluding unnecessary noise and accepting echo pulses when its frequency *bandwidth is equal to* the reciprocal of the pulse width However this adjustment should not be employed unless *it is pos* sible to assure *that the transmitted carrier* fre quency and the local oscillator are correctly tuned so that the pulses will pass through the intermedi ate-frequency amplifier The receiver bandwidth is generally determ ned ...

transmitter frequency to maintain the correct rela tionship

Pulse radar transmitter The principal opera tions within a pulse-radar transmitter are the for mation of a high power dc pulse which is then ap plied to the output tube to produce a high power carrier frequency pulse The device that provides the ...

pu ses at a high power level to the transmitter out put tube The energy output of the modulator is less than its energy input but the output pulse power exceeds the average *input* power The basic principle of operation is to charge an energy stor age circuit *slowly* from the primary supply and then discharge it rapidly into the transmitter out *put tube* A *switch is used* which rapidly changes a circuit connection and brings this about

In order to produce a dc *pulse having the de sired shape* a high voltage *transmission* line or de lay line is used as the *energy* storage circuit (*see* DELAY LINE) The input end of the line is connected to a high voltage supply but the output end is left unconnected during the *charging time A* small cur rent fl ws into the line slowly charging it to the same voltage as the primary supply When the transmission line voltage equals the supply volt age no more current flows *In order t*o deliver a h gh power pulse a switch suddenly connects the charged line to the circuit containing the transmit ter output tube The impedance of this *circuit* matches the characteristic impedance of the trans missi n l ne The charge moves out of the line *de* l vering a pulse to the output tube The rise time fall time and duration of the pulse *are determined* by the transmissi n l ne *characteristics* The trans mission line generally is formed by a *ladder net* work consisting of l mped elements The *charac* teristic impe lance *of the* transmission line is usually ... input impedance *of* the

transmitter output tube so a *transformer is* in cluded in the circuit between the two to create a *match*

The switch may be either a vacuum tube or a gas tube The voltage loss across a vacuum tube is *much* larger than across a gas tube making the vacuum tube less efficient However, a *vacuum tube can* be controlled more precisely permitting the timing of the pulse to be accurately controlled The gas tube generally a hydrogen thyratron cannot be bro ght nto cond ...

put pulse but in many radars this jitter (*which* may be a few microseconds) is not objectionable

The output *tube that* generates the carrier fre quency *pulse may* be of various types An impor tant categorization is between amplifiers and oscil lators *An* oscillator generates and determines the final frequency that is transmitted while an ampli fier increases the power of a signal already at the final frequency *The importance* of this distinction lies in the fact that some coherent radars require that the transmitted frequency and phase be closely *controlled* in which case an amplifier must be used

In the frequency band below 1000 Mc triodes and klystrons are employed At higher frequencies magnetrons klystrons and traveling wave tubes *are used* Magnetrons are used only as oscillators while the other tube types are basically *amplifiers although* they can also be connected as oscillators An *important* characteristic of traveling wave tubes is *their large* fractional bandwidth capabil ity This makes *them especially useful* for systems employing broad spectrum signals for fine range resolution *and* for systems in which the carrier fre quency is *varied See* MICROWAVE TUBE

The output *stage* is coupled to the antenna with a transmission line or wa*ve guide* Two-wire trans mission lines are generally used below 100 Mc co-axial transmission lines in the band up to 3000 Mc, and wave guides in the band above 1000 Mc *See* MICROWAVE TRANSMISSION LINES

In pulse radar systems the transmitter and re ceiver generally *share* a single antenna It is neces sary *therefore to* disconnect the receiver during the transmitter pulse to protect the *sensitive* re *ceiver* and in order not to lose any ...

... The duplexer consists *of two* switches usually gas discharge tubes which *open or close in the* presence or absence of trans mitter power During the transmitter pulse the gas *ionizes and* the impedance across the gas tube is *essentially* zero In the absence of the transmitter pulse the gas is deionized and the impedance across the tube is essentially infinite Use is made of the following impedance relations in transmis sion lines and wave guides (1) one-quarter wave-length away *from a* short circuit the input imped ance is infinite and one quarter wavelength away

Table 2 Typical measured average radar cross sections

Target	σ, m²
Small jet airplane	10
Large jet airplane	100
Small propeller airplane	150
Large propeller airplane	800
Small surfaced submarine	80
Small freighter	150
Medium freighter	8 000
Large freighter	16 000

because of maneuvers or wind buffeting. Its value can be stated only in statistical terms; it follows the Rayleigh distribution which is characterized by a single parameter. The average value of the radar cross section has been determined empirically for many types of targets, as given in Table 2.

The random fluctuation of radar cross section caused by target aspect variations is called scintillation. The scintillation power spectrum, neglecting the contribution of propellers, is approximately that of a Markov process, and possesses a half-power frequency which depends on both the target and the radar wavelength. The product of half-power frequency in cps and wavelength in cm is approximately 30 for small one- or two-seater airplanes and approximately 10 for large transports. *If the airplane has propellers, they make a large* contribution to the cross section fluctuation at discrete frequencies determined by the blade rotation rate, radar wavelength, and pulse repetition frequency.

The radar echo of ships is strongly affected by the water, which serves as a reflecting surface and leads to the production of an interference pattern. The assumption is made that the ship can be represented by a point at the same height above the water as the estimated center of reflection of the ship. Typical empirical values are given in Table 2. Reflection by the water causes the shorter radar wavelengths to be more effective for the detection of small targets on the surface.

The echo from the sea itself varies greatly with several parameters, namely, the angle of incidence of the radar wave with the ocean surface, the radar wavelength, the direction of polarization, the roughness of the sea surface, and the bearing (upwind or downwind) of the radar relative to the surface providing the echo. A smooth sea acts as a specular (without diffraction or scattering) reflector, producing very little echo in the direction of the radar. As the sea becomes rougher the waves and spray from the top of white caps cause progressively stronger echoes. The effective echoing area of the sea *surface increases* with the beamwidth of the radar and the pulse width for which the bandwidth of the receiver is matched.

Ground echoes depend upon the type of terrain under observation. In general the echoes can be resolved into a contribution due to moving objects, such as foliage in wind, and to stationary objects, such as solid ground and fixed structures. Moving objects cause scintillation of the ground echoes at a rate roughly proportional to the product of radar frequency and the object's velocity. Echoes from terrain are much stronger than those from a quiet sea, and lead to excellent radar maps along shore lines. Buildings serve as specular reflectors. At orientations which cause the radar wave to bounce between a building and the ground and then back to the radar, very strong echoes occur. Good radar maps can be obtained of cities, especially those having a coastline or rivers and bridges and a number of especially conspicuous large buildings.

Pulse radar system parameters. The performance of a pulse radar system is determined by several important parameters. The number of pulses transmitted per second is the pulse repetition frequency (prf). As the prf increases, the time between pulses decreases and the maximum range from which an echo can arrive before the next pulse is emitted decreases accordingly. In conventional radars the prf may therefore impose a limitation on maximum range. This limitation can be avoided, however, if the radar carrier frequency is changed from pulse to pulse and separate receivers are provided to accommodate the pulses that return from different ranges simultaneously.

The pulse width determines the range resolution of the system. Two targets of the same size cannot generally be separately *distinguished if* their echoes overlap. The minimum range from which the radar can receive echoes is also affected by the pulse width. The receiver cannot be brought into operation until a few microseconds after the end of the transmitter pulse because of the time required by the TR tube to deionize, as described below.

The antenna beamwidth determines the angular resolution and accuracy of the system. Two targets of the same size at the same range cannot be separately distinguished if they lie within the antenna beamwidth. This, along with the fact that the antenna gain is inversely proportional to the product of azimuth and elevation beamwidths, places a premium on using the smallest beamwidth achievable. However, the beamwidth can only be increased at the expense of an increased antenna aperture. It is difficult to construct and maintain an aperture much larger than 100 radar carrier wavelengths because of the problem of producing the required aperture illumination phase pattern. Another problem accompanying a small beamwidth is that the number of pulses per scan from each target is low. However, since the angular directivity of most radars imposes a greater limitation on accuracy than the other parameters of the system, the smallest feasible beamwidth is *used. A 1°-azimuth* beamwidth is the best generally found in search radars with an elevation beamwidth large enough to provide the desired vertical coverage.

The antenna scan speed in a search radar is chosen in accordance with data required from the system. The number of pulses per scan is an important parameter. If the radar is used for early warning, involving very long range detectio

The following examples illustrate the simplicity and convenience of radian measure

The simple formula (1) would be $s = (\pi/180)r\theta$ if degrees were used A mariner on a ship observes that a lighthouse known to be 400 ft high, subtends 0 064 rad at his eye, and writes from Eq (1)

$$400 = r(0.064) \qquad r = \frac{400}{0.064}\text{ ft} = 6250\text{ ft}$$

He then concludes that he is approximately 6200 ft from the lighthouse The very important ratios $(\sin\theta)/\theta$ and $(1-\cos\theta)/\theta^2$ approach, when θ approaches zero respective limits of 1 and ½ when radians are used but $\pi/180$ and $\pi^2/64\,800$ when degrees are used The use of radians avoids these constants throughout the many fields of application of the trigonometric functions If a particle A moving on a circle with center O and radius r has velocity v tangential acceleration a_t, normal acceleration a_n and if OA has angular velocity ω and angular acceleration α then the relations

$$v = r\omega \qquad a_t = r\alpha \qquad a_n = r\omega^2 \tag{2}$$

hold when radians are used Each of these basic formulas would involve an unwieldy constant if degrees were employed for angle measure The simplicity arising from use of the radian indicates the importance of a wise choice of basic units in mathematics and its applications. [L M K]

Radiant heating

Any system of space heating in which the heat producing means is a surface which emits heat to the surroundings by radiation rather than by con d ct on

. . .

. . . .

a es of the space to be heated See PANEL

. .

. . . .

, them appreciably but they will increase the sensible temperature of liquid or solid objects upon which they impinge

This same principle applies to all forms of radiant heating systems except that convection currents are established in enclosed spaces and a portion of the space heating is produced by convection The radiation component of convectors may be increased by providing a reflective surface on the wall side of the convector and painting the inside of the enclosure a dead black to absorb heat and transmit it through the enclosure thus increasing the temperature of that side of the convector exposed to the space to be heated

Any radiant heating system using a fluid heat conveyor may be employed as a cooling system by substituting cold water or other cold fluid This does not apply to electric radiant heating systems because at their present stage of commercial development, they are not reversible however experiments on the reversibility of thermocouples may make such a development possible in the future. See COMFORT CONTROL. [E L W]

Bibliography American Society of Heating and Air Conditioning Engineers *Heating Ventilating and Air Conditioning Guide* 1959

Radiata

Members of the Eumetazoa which have a primary radial symmetry The group includes the Coelenterata and Ctenophora although historically such groups as *Porifera* and *Echinodermata* were included

The radiate animals are in general of the tissue grade of construction and do not possess organs although simply constructed nervous digestive and muscular systems are present Incipient mesoderm is present as a mesenchyme collenchyme or mesoglea and is largely of ectodermal origin The only body cavity is the digestive cavity (gastrovascular cavity) and although primitively a simple sac it may become complex through branching and the development of compartments Typically there is a single opening into the digestive cavity, the mouth, which serves for the egestion of materials as well as for food intake

Radiata as a useful concept in the picture of the evolution of animal phyla has been questioned by J Hadzi and G Jägersten They have suggested that the Radiata are secondarily, not primarily, symmetrical radially See EUMETAZOA [C H]

Bibliography J Hadži A reconstruction of animal classification, *Syst Zool* 2 145–154, 1953, L H Hyman *The Invertebrates* vol 1, 1940, G Jägersten On the early phylogeny of the Metazoa the bilaterogasterea theory, *Zool Bidrag Uppsala* 30 321–354, 1956

Radiation

The emission and propagation of energy, also, the emitted energy itself The etymology of the word implies that the energy propagates rectilinearly and in a limited sense this holds for the many different type

ele

the there are many subdivisions

For example electromagnetic radiation, which in the most familiar energy ranges behaves in a manner usually characteristic of waves rather than of particles is classified roughly in order of decreasing wavelength as radio microwave visible ultraviolet, x rays and γ rays In the last three subdivisions, and frequently in the visible the behavior of the radiation is more particlelike than wavelike

Since the

inversely

sification

ene

fre

of

ab

ultrasonic sound can re

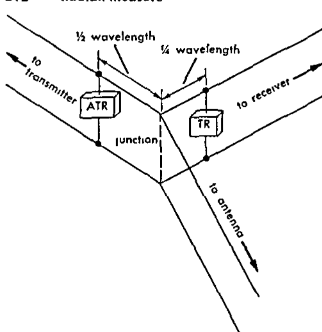


Fig 7 Schematic diagram of a duplexer The impedance across the TR and ATR devices is substantially zero during the transmitted pulse and is almost infinite at other times thereby connecting the antenna to either the transmitter or receiver respectively

from a point of infinite impedance the input impedance is zero, and (2) one half wavelength away from a point of infinite impedance the input impedance is infinite The transmission lines or wave guides from the transmitter output stage and the receiver input stage are brought to a junction and the common point is connected to the antenna A gas tube (TR) is located in parallel with the line connected to the receiver one-quarter wavelength from the junction and another gas tube (ATR) is placed in series with the line connected to the transmitter one half wavelength from the junction This arrangement is shown in Fig 7

When the transmitter pulse is on both gas tubes ionize The one in series with the line from the transmitter forms part of the conductive path carrying the output to the antenna But the gas tube across the line leading to the receiver produces a short circuit across the line and presents an infinite impedance in the direction of the receiver at the junction Therefore none of the transmitter power goes to the receiver

In the absence of the transmitter pulse both tubes act as open circuits The tube in series with the transmitter line causes the impedance looking toward the transmitter from the junction to be infinite so none of the echo power flows to the transmitter The gas tube across the receiver line causes the impedance looking toward the receiver from the junction to be zero so all the echo power from the antenna is channelled to the receiver

The gas tube located in the receiver line is called the TR (transmit receive) tube and the one in the transmitter line is called the ATR (anti transmit receive) tube Spark gaps or vacuum tubes are sometimes employed If a transmitter's average power output is extremely large 100 kw or more, the life of a gas ATR tube may be so short that a spark gap must be used instead On the other hand the instantaneous pulse power of a transmitter may be too small (below 100 watts) to fire a gas TR tube satisfactorily, in which case a grid controlled vacuum tube is used in its place with a signal applied to its control grid to decrease the vacuum tube's resistance at the appropriate instant See GAS TUBE [R J B]

Bibliography J F Reintjes and G T Coate, *Principles of Radar*, 3d ed 1952, L N Ridenour (ed), *MIT Radiation Laboratory Series* 28 vols, 1947–1953, A T Starr, *Radio and Radar Techniques*, 1953

Radian measure

A radian is the angle subtended at the center of a circle by an arc of the circle equal in length to its radius It is proved in geometry that equal central angles of two circles subtend arcs proportional to their radii, and conversely Hence the radian is independent of the length of the radius The figure represents two circles of radius r Arc AB of length r subtends 1 radian (rad) at the center O of the circle, and arc $A'B'$ of length s subtends θ rad at its center Since arcs on equal circles are proportional to their subtended central angles $s/r = \theta/1$ or

$$s = r\theta \tag{1}$$

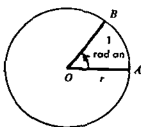
If $\theta = 2\pi$, $s = 2\pi r$ the circumference of the circle Therefore 2π rad is the complete angle about a point or 360° and

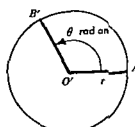
$$2\pi \text{ rad} = 360°$$
$$1 \text{ rad} = 360°/2\pi = 57\,2958° = 57°\,17'45''-$$
$$1° = 2\pi/360 \text{ rad} = 0\,0174533+ \text{ rad}$$

Observe that

$$30° = 30\pi/180 \text{ rad} = \pi/6 \text{ rad}$$
$$45° = \pi/4 \text{ rad}$$
$$60° = \pi/3 \text{ rad}$$
$$90° = \pi/2 \text{ rad}$$
$$135° = 3\pi/4 \text{ rad}$$

The degree as a unit of angle has come down from antiquity However its use in various theories involves clumsy constants The use of the radian avoids these constants The radian is employed generally as a measure of angle in theoretical discussions when no unit of angle is mentioned the radian is understood





Radian measure

instrument Special glass or fused quartz, transparent to most of the sun's radiation is used for the envelope

Net radiation measurement An instrument which measures the net effect of all the upward and downward solar and terrestrial radiation currents must be equally sensitive to the whole range of wavelengths mentioned earlier No window material completely transparent over this wide range exists however some film plastics approach complete transparency In better instruments, the detector a horizontal plate with provision to measure the difference in temperature between upper and

loss is held constant by a jet of air from an electric blower {V E S}

Radiation biochemistry

and organisms and with the balance and integrated metabolic reactions in which these take part (see

widely used in medicine for diagnostic and therapeutic purposes but the applications of radiation in industry have increased For example the use of luminous paint on instrument dials the hazards connected with the development of atomic energy and the possible use of atomic weapons in war make research

bre of defensive barriers puts ionizing radiations in a unique position as compared to other noxious agents which are limited in penetration or selectively active on special cell constituents

The cells of living matter consist of a cell membrane surrounding protoplasm

The of the cell is inhomogeneous highly organized and equipped with a series of enzymes which make complicated metabolic reactions possible See CELL (BIOLOGICAL)

The indirect and the direct modes of action of radiation have been proposed as mechanisms of radiation effects These modes of action are discussed in the following paragraphs

Indirect action The water content of cells is about four times greater than their dry weight The effect of radiation on the water has consequences for the solid matter contained in it

When water is irradiated it is split into the primary radiation products hydroxyl (OH) radicals and hydrogen (H) atoms which are highly reactive uncharged chemical entities Oxidation and reduction reactions are brought about when the primary radiation products collide with solutes the substances dissolved in water In the absence of solutes however, they quickly recombine to form water The irradiation thus acts indirectly on solutes via the water Consequences of this indirect action are the dilution effect the protection effect and the oxygen effect

Dilution effect This effect causes a dilute solution to appear more sensitive to radiation than a concentrated one because a given dose of radiation

solute molecules chemically changed will be large if the solution is dilute but small if it is concentrated

Protection effect This comes into play when there are two or more solutes in solution which are capable of

will compet

fore fewer

cies of solute than would be the case if only these solutes were present in the solution In other words, there is a mutual diminution of the radiation effect and each solute appears protected by the others against radiation This protection effect will vary in accordance with the absolute amounts of solutes and with their specific capability of reacting with radicals that is with their capability of acting as acceptors

this solute can still react with radicals Before radiation starts the solution of an enzyme (catalyst of living matter) contains the dissolved active enzyme

active molecules then represent a second solute which competes with the still active molecules for radicals The result is that equal but subsequent increments of radiation become less and less effective so that an activity dose curve takes on an exponential shape

namely to HO_2 radicals and to the stable product hydrogen peroxide (H_2O_2) These as oxidative agents, can react with solutes and thereby enhance the radiation effect when compared

decomposition of solutes after irradiation has ceased The interference of hydrogen peroxide complicates the clarification of reaction mechanisms It has been found that nitric oxide increases

sult, for example, from explosions or other sources so loud that exceptional waves are set up because the large amplitudes of the source vibrations exceed the elastic limit of the transmitting medium Ultrasonic sound can be produced by means of crystals which vibrate rapidly in response to alternating electric voltages applied to them There is a nearly infinite variety of sources in the sonic range *See* SOUND

Particle radiation occurs in a large number of types, the traditional ones being the α and β rays of radioactivity In addition, cosmic rays consist of protons, neutrons, and heavier nuclei, along with β and γ rays, mesons, and the so called strange particles *See* COSMIC RAYS, ELEMENTARY PARTICLE

[M H H]

Radiation, terrestrial

Electromagnetic radiation originating from the earth and its atmosphere at wavelengths determined by their temperature It is sometimes called thermal radiation The units are energy per unit area per unit time, such as cal/(cm^2) (sec) The atmosphere emits, absorbs, and transmits radiation and the net flux of radiation at any point depends upon the distribution with height of temperature and water vapor Heating and cooling due to the vertical divergence of terrestrial radiation provide a major part of the potential energy changes necessary to drive the atmospheric wind system (*see* HEAT BALANCE TERRESTRIAL ATMOSPHERIC) Terrestrial radiation is also responsible for maintaining the air temperature near the ground within limits necessary for comfortable living *See* GREENHOUSE EFFECT, TERRESTRIAL

At terrestrial temperatures practically all emission of radiation is at wavelengths greater than 4 μ (1 μ = 10^{-4} cm), that is within the infrared part of the spectrum The earth's surface and all but the thinnest clouds emit, at all wavelengths, radiation of intensity only slightly less than that of black body radiation corresponding to their temperatures, and absorb almost all the infrared radiation that reaches them *See* ABSORPTION (ELECTROMAGNETIC RADIATION) The absorption and emission by the atmosphere apart from clouds is, on the other hand, selective, and depends on the spectral position and intensity of the rotation vibration and pure rotation bands of its polyatomic molecules (*see* BAND SPECTRUM, OPTICS) The minor gases of the atmosphere have numerous absorption bands in the infrared—water vapor at 5–8 and beyond 15 μ, carbon dioxide at 13–17 μ, and ozone at 9–10 and 14 μ Within these bands the absorption coefficient, or absorption per unit mass of absorber, has many maxima, each maximum marking the position of a spectral line These lines have finite

the transition of the molecule from one quantum state to another by absorption or emission of a photon of frequency corresponding to the line frequency The line shape, or variation of absorption coefficient with frequency, is determined by molecular collisions in the lower atmosphere and by random thermal motions in the upper atmosphere The absorption coefficient is also a function of temperature and pressure and because of the marked spatial variation of these parameters has itself a considerable variation, especially in the vertical

[L D K]

Radiation-measuring devices These instruments are used to obtain the radiant energy exchange, solar and terrestrial, between the sun, space, and the earth The energy covers a broad range of wavelengths from about 0 17 to 100 μ Solar radiation ranges from 0 17 to 4 μ with the maximum at 0 49 μ, whereas terrestrial (earth) radiation ranges from about 3 to 100 μ with the *maximum typically at 10 μ It is important* that the sensitivities of these instruments be independent of the wavelength of the incident radiation. Because of this requirement practically all radiation measuring instruments in meteorological use first convert the radiant energy into heat by absorption on a blackened target The actual measurement is made on the resulting heat flow

Energy in the solar beam This is measured with a normal incidence pyrheliometer A number of instruments of this type are in use, however, two are considered secondary standards The Abbot silver disk pyrheliometer consists of a thermally insulated blackened silver disk mounted at the end of an open end tube The tube has a shutter and baffles to collimate the beam During a measurement the disk is alternately exposed to and shaded from the sun's radiation for equal periods of 2 min A mercury thermometer imbedded in the silver disk indicates the temperature change during the exposed and shaded periods This, together with the heat capacity of the disk and other constants, is used to calculate the radiation The Angstrom compensation pyrheliometer has two nearly identical targets equipped with electric heaters at the bottom of a similar tube A sunshade is arranged so one target is exposed to the sun while the other is shaded Electric current is supplied to the shaded target until its temperature, determined by thermocouples, is equal to the sunlit one The power required is a measure of the solar radiation for the target area

Total sun and sky The total solar energy from sun and sky striking a horizontal surface is meas-

are of

sulated

horizontal target within a partially evacuated transparent sphere or hemisphere to protect it from convective and conductive heat losses Part of the *target* is a ring coated with *dull black* to absorb as much sunlight as possible, the remainder, a disk and a ring, is coated with white to reflect as much sunlight as possible A thermopile measures the temperature difference between the white and black portions of the target and is the output of the

instrument. Special glass or fused quartz, transparent to most of the sun's radiation is used for the envelope

Net radiation measurement An instrument which measures the net effect of all the upward and downward solar and terrestrial radiation currents must be equally sensitive to the whole range of wavelengths mentioned earlier No window material completely transparent over this wide range exists however ome film plastics approach complete transparency In better instruments the detector a horizontal plate with provision to measure the difference in temperature between upper and lower surfaces, is exposed without benefit of a wind screen In the Albrecht net *radiometer* the effect of the wind is measured by adding a known quan of heat electrically to one of two identical tes In the Gier and Dunkle instrument the wind s is held constant by a jet of air from an electric wer [V E S]

adiation biochemistry

... ty of the building stones of living tissues d organisms and with the balance and integrated etabolic reactions in which these take part (*see*

... idely used in medicine for diagnostic and thera eutic purposes but the applications of *radiation* n industry have increased For example the use of uminous paint on instrument dials, the hazards connected with the development of atomic energy and the possible use of *atomic weapons in war* make research in the effects of radiation important

The capability of penetrating to every part of the interior of cells without being obstructed by membranes or defensive barriers puts ionizing radiations in a unique position as compared to other noxious agents which are limited ...

The ... of the cell is inhomogeneous highly organized and equipped with a series of enzymes which make complicated metabolic reactions possible See CELL (BIOLOGICAL)

The indirect and the direct modes of action of radiation have been proposed as mechanisms of radiation effects These modes of action are discussed in the following *paragraphs*

Indirect action The water content of cells is about four times greater than their *dry weight* The effect of radiation on the water has consequences for the solid matter contained in it.

When ... the primary rad ... icals and hydrogen ... reactive uncharged chemical entities Oxidation and reduction reactions are brought about when the primary radiation products collide with solutes the substances dissolved in water In the absence of solutes however, they quickly recombine to form water The irradiation thus acts indirectly on solutes via the water Consequences of this indirect action are the dilution effect, the protection effect and the oxygen effect.

Dilution effect This effect causes a dilute solution to appear more sensitive to radiation than a concentrated one because a given dose of radiation results in the formation of a given number of radicals which will react with a corresponding number of solute molecules Therefore the *proportion* of solute molecules chemically changed will be large if the solution is dilute, but small if it is concentrated

Protection effect This comes into play when there are two or more ...

... ties of solute than would be the case if only these solutes were present in the solution In other words there is a mutual ... and ... again ... accordance with the *absolute* amounts of solutes and with *their* specific capability of reacting with radicals that is with their capability of acting as acceptors for radicals

The protection effect operates also in a one-solute solution if the irradiation products formed from ...

... cals The inactive molecules then represent a second solute which competes with the still active molecules for radicals The result is that equal but subsequent increments of radiation become less and less effective so that an activity-dose curve takes on an exponential shape *as shown in the diagram*

Oxygen effect The primary OH radicals and H atoms give rise to other radicals if oxygen is present in solution namely to HO_2 radicals and to the stable product hydrogen peroxide (H_2O_2) These, as oxidative agents can react with solutes and thereby enhance the radiation effect when compared with oxygen free solutions, hence the term oxygen effect Hydrogen peroxide may also lead to the formation of organic peroxides which can be responsible for an after effect, *that is*, a continuation of decomposition of solutes after irradiation has ceased The interference of hydrogen peroxide complicates the clarification of reaction mechanisms It has been found that nitric oxide increases

sult, for example, from explosions or other sources so loud that exceptional waves are set up because the large amplitudes of the source vibrations exceed the elastic limit of the transmitting medium Ultrasonic sound can be produced by means of crystals which vibrate rapidly in response to alternating electric voltages applied to them There is a nearly infinite variety of sources in the sonic range *See* SOUND

Particle radiation occurs in a large number of types the traditional ones being the α and β rays of radioactivity In addition, cosmic rays consist of protons, neutrons, and heavier nuclei, along with β and γ rays, mesons, and the so called strange particles *See* COSMIC RAYS, ELEMENTARY PARTICLE [M H H]

Radiation, terrestrial

Electromagnetic radiation originating from the earth and its atmosphere at wavelengths determined by their temperature It is sometimes called thermal radiation The units are energy per unit area per unit time, such as cal/(cm^2) (sec) The atmosphere emits, absorbs, and transmits radiation, and the net flux of radiation at any point depends upon the distribution with height of temperature and water vapor Heating and cooling due to the vertical divergence of terrestrial radiation provide a major part of the potential energy changes necessary to drive the atmospheric wind system (*see* HEAT BALANCE, TERRESTRIAL ATMOSPHERIC) Terrestrial radiation is also responsible for maintaining the air temperature near the ground within limits necessary for comfortable living *See* GREENHOUSE EFFECT, TERRESTRIAL

At terrestrial temperatures practically all emission of radiation is at wavelengths greater than 4 μ (1 μ = 10^{-4} cm), that is, within the infrared part of the spectrum The earth's surface and all but the thinnest clouds emit, at all wavelengths, radiation of intensity only slightly less than that of black body radiation corresponding to their temperatures, and absorb almost all the infrared radiation that reaches them *See* ABSORPTION (ELECTROMAGNETIC RADIATION) The absorption and emission by the atmosphere apart from clouds is on the other hand, selective, and depends on the spectral position and intensity of the rotation vibration and pure rotation bands of its polyatomic molecules (*see* BAND SPECTRUM, OPTICS) The minor gases of the atmosphere have numerous absorption bands in the infrared—water vapor at 5–8 and beyond 15 μ, carbon dioxide at 13–17 μ, and ozone at 9–10 and 14 μ Within these bands the absorption coefficient, or absorption per unit mass of absorber, [illegible]

of atmospheric bands is of considerable importance for radiative heat transfer Each line is the result of the transition of the molecule from one quantum state to another by absorption or emission of a photon of frequency corresponding to the line frequency The line shape, or variation of absorption coefficient with frequency, is determined by molecular collisions in the lower atmosphere and by random thermal motions in the upper atmosphere The absorption coefficient is also a function of temperature and pressure, and because of the marked spatial variation of these parameters has itself a considerable variation, especially in the vertical [L D K]

Radiation measuring devices These instruments are used to obtain the radiant energy exchange, solar and terrestrial between the sun space, and the earth The energy covers a broad range of wavelengths from about 0 17 to 100 μ. Solar radiation ranges from 0 17 to 4 μ with the maximum at 0 49 μ, whereas terrestrial (earth) radiation ranges from about 3 to 100 μ with the maximum typically at 10 μ It is important that the sensitivities of these instruments be independent of the wavelength of the incident radiation. Because of this requirement, practically all radiation measuring instruments in meteorological use first convert the radiant energy into heat by absorption on a blackened target The actual measurement is made on the resulting heat flow

Energy in the solar beam This is measured with a normal incidence pyrheliometer A number of instruments of this type are in use, however two are considered secondary standards The Abbot silver disk pyrheliometer consists of a thermally insulated blackened silver disk mounted at the end of an open end tube The tube has a shutter and baffles to collimate the beam During a measurement the disk is alternately exposed to and shaded from the sun's radiation for equal periods of 2 min A mercury thermometer imbedded in the silver disk indicates the temperature change during the exposed and shaded periods This, together with the heat capacity of the disk and other constants, is used to calculate the radiation The Angstrom compensation pyrheliometer has two nearly identical targets equipped with electric heaters at the bottom of a similar tube A sunshade is arranged so one target is exposed to the sun while the other is shaded Electric current is supplied to the shaded target until its temperature determined by thermocouples is equal to the sunlit one The power required is a measure of the solar radiation for the target area

Total sun and sky The total solar energy from sun and sky striking a horizontal surface is measured with a hemispherical pyrheliometer The Kimball Epply and Moll Garczynski instruments are of this type They consist of a thermally insulated horizontal target within a partially evacuated transparent sphere or hemisphere to protect it from convective and conductive heat losses. Part of the target is a ring coated with dull black to absorb as much sunlight as possible, the remainder, a disk and a ring, is coated with white to reflect as much sunlight as possible A thermopile measures the temperature difference between the white and black portions of the target nd is the ou [illegible]

these values both oxidation and reduction are pos sible The redox potentials of most redox systems in cells are in the oxidation range but will display different resistance to oxidative changes according to their total concentration as well as to the re spective ratio of their reduced state to their oxi dized state Biological redox systems are normal constituents in cells and form the link in many met abolic steps There is opportunity therefore for in terference with the normal metabolism Some in stances of biochemical redox systems are cysteine to cystine sulfhydryl to disulfide (SH—SS) gluta thione prosthetic groups of enzymes such as flavo protein coenzymes I and II ascorbic acid and many others *See* BIOLOGICAL OXIDATION

Redox systems may also occur in which the re versibility of the reaction is impaired If the oxi dized or the reduced state of the reacting compound suffers secondary changes such as the formation of a polymer or if the reduced form is capable of forming a molecular compound with the oxidized form no proper *equilibrium* will be established

the main chain can form cross linkages between neighboring chains In solution they form finely dis persed colloidal systems The variety of existing proteins is very great but all proteins have one re action in common namely the property of denatur ation This is usually an irreversible change which occurs when proteins combine with certain chemi cals or when heat or radiation are applied

The *denaturation* manifests itself as coagulation with subsequent insolubility It has been found that ionizing radiations lower the resistance of proteins to thermal denaturation After irradiation protein

established that the oxidative attack by OH radicals is directed towards the peptide link age This results in the formation of high molecular weight carbonyl (C—O) compounds and of keto acids and the release of ammonia The radiation

neighboring chains Oxidation can occur even fur ther beyond the S-S stage without denaturation or marked instability of the protein It has further been observed that after hydrolysis of irradiated serum albumin several amino acids are partly de stroyed Examination of irradiated protein solu tions by spectrophotometry also reveals changes Bovine serum albumin serum globulin and egg al bumen show an increase in optical density wh ch appa

tyros

tain

Not only is protein as such changed by radia tion but its building stones the amino acids are affected *See* SPECTROPHOTOMETRIC ANALYSIS

Amino acids The principal effect is the loss of ammonia from or deamination of the amino acids The extent of deamination varies with experimen tal conditions but an important point is that it also varies with the chemical configuration of the amino acid itself If the amino group is in the alpha (α) position attached to the carbon atom next to the carboxyl group (COOH), as in α alanine $CH_3 CH(NH_2) COOH$ the loss of ammonia is nearly twice as great as for β alanine $CH_2(NH_2) CH_2 COOH$ In this compound the *amino group is at* tached to second carbon atom that is in the β po sition This is an example of the specificity of ra diation effects More ammonia is split off from his tidine where the glyoxaline part may contribute to the yield

In experiments in which radiation products other than ammonia were examined it was found that alanine irradiated in a vacuum *yielded acetalde*

acid and carbon dioxide It is however claimed that the appearance of the various products depends on whether radiation doses applied have been mod erate or massive This claim is made because some of the radiation products are not due to initial re action but are formed by further oxidation or de carboxylation of glyoxylic acid *See* AMINO ACIDS

Enzymes These are proteins which differ from other proteins in their ability to act as catalysts Enzymes speed up specific chemical reactions They either have special active groups in their make up enabling them to combine with their specific sub strates on which they act, or they are more or less firmly linked to a nonprotein partner or prosthetic group which in cooperation with the protein part functions as a highly specific catalytic system

As far as their protein nature is concerned en zymes will undergo the same general changes as de scribed for other proteins with consequent loss of activity There are some enzymes the S-H enzymes in which that group is essential for enzymic activ ity This S-H group is particularly sensitive to ra diation and may

seat

If th

far

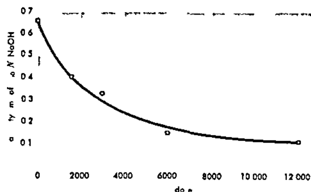
addition of the tripeptide glutathione containing S-H *See* ENZYME

An example of an enzyme containing a prosthetic group is D-amino acid oxidase which specifically oxidizes D-amino acids only This enzyme can be split into flavin adenine dinucleotide a nonprotein and a specific protein Neither part on its own has enzymic activity but when combined they con stitute the complete act

che

sho

con



Exponential relation between degree of inactivation and x ray dose for carboxypept dase solut ons (W M Dale 1940 *reproduced by permission of the ed tors of the Biochemical Journal*)

sensitivity to radiation under conditions of anoxia to an extent similar to that observed with oxygen The explanation put forward for this effect is that both nitric oxide and oxygen have an equal affinity for carbon radicals

Direct action Since the dissipation of radiation energy is not confined to the solvent direct ionization with subsequent chemical change will occur in solute molecules themselves The frequency of such an event (single hit) increases in proportion to the concentration of the solution and reaches its maximum when one hit is scored per molecule of a dry substance

Some investigators believe that the nucleus of a cell contains a vital and sensitive structure in which the primary event that is an ionization has to occur or that an ionizing particle has to pass near or through it in order to cause a chemical alteration that results in the biological effect subsequently observed This is the target hit theory

An important practical application of the direct hit theory is the determination of the molecular weight—sometimes even the shape of large biologically active molecules irradiated in the dry state The underlying assumption is that the destruction of the biological function for example of enzymic activity or of the ability of a virus to infect is caused by a primary ionization produced by a fast charged particle passing through the molecule It is possible to calculate the radiation dose which will produce on the average one ionization per molecule From the number of ionizations occurring per unit volume the target size can be assessed This method has been successful in a number of cases

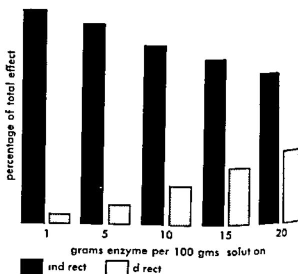
Direct plus indirect action The share taken by the direct and indirect modes of action is illustrated in the diagram which shows the inactivation of an aqueous solution of an enzyme (carboxypeptidase from pancreas) at various concentrations

Since water is always in excess in living tissues except in bone structure and fatty tissue the opportunity of reactions with radicals is always preponderant A distinction between the direct and indirect modes of action becomes increasingly trivial as the source of active radicals approaches the molecule or structure upon which they act As a result the two modes of action merge in the immediate vicinity of the target constituting a direct hit

Measurement of radiation sensitivity The problem of a more detailed analysis of the radiation products from substances of biochemical importance has been solved in only a few instances However the protection effect has made it possible to assess over all radiation sensitivity over a wide range of substances extending from small molecules to the large molecules of proteins From the degree of inactivation of an enzyme solution of known concentration in the absence and in the presence of a protective substance a value of the reactivity with radicals of the substance in question can be derived In this way it has been found that specificity of radiation effect can be detected in small molecules Thus it has been shown that sulfur in organic molecules causes a high degree of radiation sensitivity In general specific effects are not discernible in large molecules because of the great number of their reactive groups but it can be said that the protective power of large molecules is approximately proportional to their molecular weight

Effect of radiation The effect of radiation upon biological material is discussed in the following paragraphs

Biological reduction oxidation systems Most reactions of solutes with radicals consist of oxidations and reductions and it has been proposed that the OH radicals and H atoms in irradiated water constitute a redox system having an equivalent redox potential (ERP) This system reacts with a range of redox systems as solutes in such a way that for redox potentials greater than −0 52 volts oxidation occurs and for potentials less than −1 1 volts only reduction takes place This depends on such things as the type of radiation pH presence or absence of oxygen and so forth In the range between



Relat ve contributions of indirect and d rect act ons to the total effect of x rays on carboxypeptidase in solut on (W M Dale 1947 *reproduced by perm ss on of the ed tors of the Br t sh Journal of Rad ology*)

In a comparative study of the effect of the densely ionizing alpha radiation versus x radiation, on the enzyme carboxypeptidase in solution, it was found that alpha radiation was only one twentieth as effective as x radiation The effect appeared to be entirely due to the delta rays, which branch off the alpha ray track as spurs and which have a similar ion density to x rays, and not to the primary ionization column of the alpha ray track Quite generally, the lower efficiency of alpha radiation on substances in aqueous solution is in contrast to its higher efficiency on biological systems, such as the breakage of chromosomes in cells See RADIATION CYTOLOGY

Nucleic acid and nucleoproteins These are the important chemical building stones of the genetic material contained in the chromosomes of cell nuclei The nucleoproteins are saltlike unions of a nucleic acid with basic proteins such as protamine or a histone Two types of nucleic acids exist, ribonucleic acid (RNA) and deoxyribonucleic acid (DNA), both of which are built up from nucleotides containing a nitrogenous base, a pentose sugar, and phosphoric acid The base forms an ester with the phosphoric acid Both nucleic acids contain the bases adenine, guanine, and cytosine but RNA has uracil and the sugar D ribose whereas DNA contains thymine and the sugar D deoxyribose RNA is formed predominately in the cytoplasm and DNA in the nucleus Since the molecular weight is of the order of $6–8 \times 10^6$, each molecule must contain a great number of nucleotide units The structure of nucleic acid has been proposed to be two complementary, helical, nucleotide threads, joined together by hydrogen bonds between the basic guanine and cytosine and between adenine and thymine The nucleic acids are similar in organization to protein, with its amino acid units and its cross linkages Ir radiation breaks down the hydrogen bonds between the DNA threads and denaturation by heat is facilitated after irradiation The instability of nucleic acids is evident from the short heating (15 minutes in boiling water) required for a marked decrease in viscosity of DNA accompanied by decrease in molecular weight A similar decrease of viscosity in dilute DNA solution is effected by relatively small doses of radiation

Chemical changes require large doses of radiation Among the reactions observed are deamination, liberation of free purine decrease in optical density, increase in amino nitrogen, breakage of the pyrimidine ring, oxidation of the sugar moiety (ribose portion), and liberation of some inorganic phosphate The conditioning effect of radiation is shown by the fact that acid hydrolysis subsequent to irradiation, liberates free phosphate more quickly than would have occurred without irradiation In the irradiation of the breakdown products of nucleic acids, such as nucleotides nucleosides, and the purine and pyrimidine bases, the radiation effects resemble those from nucleic acids [W M D]

Radiation biology

A study of the influence of radiation due to light or ionizing radiation, such as x rays or fast particles, on living systems Radiation biology is a very broad subject, ranging from a consideration of the effects of visible light on metabolism to the effects of cosmic rays on whole organisms Because of the breadth of the subject, it is studied, to a large extent, in separate areas and the first large division into areas separates the fields of ionizing radiation effects and photon effects The term photon effects means the action of ultraviolet, visible, and infrared light in the region where ionization does not occur See RADIATION, X RAY(S), PHYSICAL NATURE OF

Ionizing radiation effects The process of ionization is characterized not only by the release of an electron from an atom, with the formation of a positive residue (the whole being an ion pair) but also by the quite large amount of energy associated with the process A typical chemical bond has an energy of 3 electron volts (ev), a typical primary ionization an energy of 100 ev or 33 times as large In consequence, ionizing radiation exerts a powerful molecular action Such action is not very specific, it occurs anywhere at random, and the ionizations are relatively far apart. Ionizing radiation therefore produces random energy releases of great size and hence generally great disruptive effect. Such disruptive effect may be on functioning units of the organism when it is termed direct action or it may be on the water moiety, when active radicals, notably OH and H are formed These exert a gentler yet potent action which can occur at some distance from the original ionization, because of radical diffusion This is termed indirect action.

Dose of ionizing radiation To measure any effect of ionizing radiation the radiation must be measured in amount Such measurement is called dosimetry and the amount given the dose Three units are in use the roentgen (r) defined as that radiation which will release one electrostatic unit of separated charge in 0 001293 gram of dry air, the rad, which is defined as that amount of radiation which will release 100 ergs in 1 gram of representative biological tissue and the roentgen equivalent physical or rep which is meant to be an energy defined unit equivalent to 1 r and which approximates 93 ergs/gram of tissue

Some effects of ionizing radiation A huge variety of effects exist. A brief sampling of these is given

1 Survival after whole body irradiation A whole animal, such as a mouse which has been irradiated, shows no marked immediate effect If the animal is followed for a few days it is found that a definite increase in mortality occurs above 400 r with a mean lethal dose at 500 r and a rapid increase in mortality rising to 100% at 1000 r Such a dose survival curve is called sigmoid In such a process death is due to effects on a variety of organs the blood forming organs d any rapi [illegible] viding tissue being the

2 Pr ... mutations Ionizing radiation produce ... the pos ... such m...

3 *Production* of chromosome breaks ... of chromosome abnormalities including chromosome breaks are produced by ionizing radiation Such breaks seem to require of the *order of* 50 ion pairs *to achieve* a break Chromosome breaks can restitute and probably the majority do so In the presence of dissolved oxygen the rate of chromo- ... factor of ...

4. Delay of cell *division* Ionizing radiation inhibits cell division in a marked degree The degree of delay increases with the dose

5 Formation of giant cells Cells in which division is delayed may *grow into* giant cells *many* times the normal size in the case of mammalian cells or into long filaments in the case of bacteria

6 Reduction of survival If means for studying the ability of a cell to divide are available it is found that animal plant and bacterial cells and viruses are reduced in survival For human cells the survival follows a nearly exponential course obeying a roughly two-hit relation corresponding ...

single hit, or simply exponential

7 Action on biological macromolecules. All important biological macromolecules—DNA RNA enzymes and antigens—are destructively affected by ionizing *radiation Roughly speaking if one* ionization occurs within the molecule or if one active radical in the case of nucleic acid or one to ten radicals in the case of protein reach the molecule it loses its *function Such actions are affected* by oxygen tension in general The process in the case of nucleic acid appears to be breaking or cross linking and in the case of protein the opening of S-S bonds

8 Action on metabolism and protein synthesis Metabolism is reduced by ionizing radiation but *much less* sensitively than division or the formation of lethal mutations The *process of protein* synthesis is reduced as is the formation of microsomal particles or ribosomes which are agents in *protein synthesis*

Action of ultraviolet light Ultraviolet light differs from ionizing radiation in that it is less energetic and also much more specific For any effect of ultraviolet light *to take* place it must first be absorbed a process which is wavelength *dependent* The most significant absorption of ultraviolet light is by *nucleic* acids which have a broad absorption maximum at 2600 A The absorption and action is predominantly in the pyrimidines of nucleic acid and the chemical result may be the addition of H and OH to the pyrimidine, thus altering it chemically Many photons have *to be absorbed* in order for an *effect* to be observed, and the ratio of changes produced to photons absorbed is called the quantum yield. Quantum yields range from 0 01 for molecules to 10^{5} for viruses

There is also *absorption in protein,* notably by *aromatic amino* acids and cystine Although the cystine absorption is neither great nor specific, it seems to account for the action of ultraviolet light on proteins

Because of *the need for* absorption, ultraviolet light rarely exerts lethal action on whole animals. The effects of ultraviolet radiation are discussed in the following paragraphs

1 Burns *Ultraviolet light produces drastic action on the skin* familiar to all as sunburn

2 Production of mutations Ultraviolet light produces mutation in all cells including bacteria and viruses. It probably does so by relatively small chemical *alteration in the DNA*

3 Production of chromosome breaks Ultraviolet light produces chromosome breaks, though of a less drastic kind than ionizing radiation, being more commonly in one of two chromatids rather than the *whole chromosome*

4 Delay of cell division and giant cell formation Cell division is inhibited by ultraviolet light and such cells can grow to giant size

5 *Reduction of survival Ultraviolet light* reduces survival though not so simply, in general as for ionizing radiation, there being a tendency for a fraction to show less sensitivity The multiple hit character is very marked in some cases

6. *Photoreactivation If cells or virus-infected* cells are subjected to illumination by light in the blue or near ultraviolet range, the degree of survival is markedly increased a phenomenon known as *photoreactivation*

7 Action on metabolism and protein synthesis These are affected by ultraviolet light though not in just the same way as for ionizing radiation More effect takes place on the DNA synthetic process, which is less affected by ionizing *radiation*

Theory of radiation biology. At present, the statement can be made that ionizing radiation effects chromosome damage and inactivates macromolecules while *ultraviolet* light produces chemical change in DNA, but as yet there is no well substantiated theory of radiation biology and such a theory is clearly badly needed *See* LINEAR ENERGY TRANSFER (BIOLOGY) RADIATION BIOCHEMISTRY, RADIATION CYTOLOGY, RADIATION INJURY (BIOLOGY), RADIATION MICROBIOLOGY [E C P O]

Bibliography Z M Bacq and P Alexander, *Fundamentals of Radiobiology,* 1955, M Forssberg and A Errera (eds) *Mechanisms in Radiobiology,* 1960 A Hollaender (ed), *Radiation Biology,* 1954, D E Lea *Actions of Radiations on Living Cells,* 2d ed 1955

Radiation chemistry

The chemistry of the effects of high energy radiation on matter High *energy* radiation includes the emanations associated with radioactivity and fis-

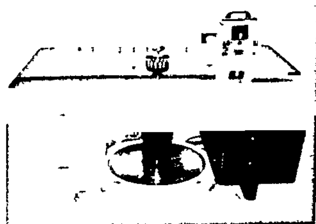
sion (that is, helium nuclei, electrons, γ rays, and neutrons), associated atom and fission recoils, and the artificial analogs of such emanations, such as accelerated electrons, protons, deuterons, helium nuclei, carbon nuclei, and x rays. Physicists tend to group all effects of radiation under the term radiation damage.

Sources of high-energy radiations in the laboratory include radioactive nuclides (for example, Co^{60} or H^3), x-ray tubes, Van de Graaff Generators, the betatron, the cyclotron, and similar instruments. A Co^{60} source is shown in the photograph.

Energy transfer. Energy is transmitted to the irradiated material by momentum transfer and by excitation and ionization. The latter two always accompany the former. Momentum transfer is characteristic of processes involving neutrons; it is always involved to some extent in particle effects and is an important contributor to heavy-particle (for example, proton) effects.

In a momentum-transfer interaction, the usual effect is the ejection of a nucleus from its molecular or crystalline structure; this is known as the Wigner effect. In the case of crystalline material the process is called discomposition. Discomposition results not only directly from neutron impact but also secondarily from impacts involving high-energy displaced nuclei. In crystalline elements, momentum-transfer effects are detected by changes in electrical and thermal conductivities, elastic moduli, and dimensions. In crystalline compounds, effects include chemical changes, electron trapping, color production, and so on.

Chemical yields are expressed in the older literature as ion pair yield M/N, which is the number of molecules converted or produced per ion pair initially produced by the radiation. The modern literature uses the 100-ev yield G, which is the number of molecules converted or produced per 100 ev energy input. The term M/N is now used only in



An inexpensive earth-shielded structure employed at Notre Dame University for exposure of samples to Co^{60} γ-radiation. Samples are lowered into a central tube. The cobalt is arranged in tubes around the central tube about 7 ft below ground. About 1200 curies of Co^{60} were originally introduced at the bottom level. As the Co^{60} decays, it may be replaced or more can be added.

cases where N, the number of ion pairs, is actually determinable from experimental data. A convenient rule of thumb for reading the older literature is $G \cong 3M/N$. Yields range from values such as $G(C_6D_6 \rightarrow D_2) = 0.0113$ to those such as

$$G(H_2 + Cl_2 \rightarrow 2HCl) \sim 10^5$$

Energy input may be determined directly from current and voltage with machine sources or indirectly by chemical dosimetry. The Fricke dosimeter (acidic ferrous sulfate solution) is such a secondary standard, for high-energy γ and x rays $G(Fe^{2+} \rightarrow Fe^{3+}) \cong 15.6$.

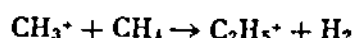
Theoretical considerations regarding primary physical processes indicate that for other than momentum transfer it is sufficient to address attention exclusively to the role of fast moving charged particles. A 1-Mev charged particle, unlike a parent photon which produces only one ionization (as in a Compton process) may produce a total of about 10^5 ions (and electrons) and excited molecules. The distribution of such primarily produced entities is affected greatly by the nature of the radiation and by the state of aggregation of the material irradiated. In condensed systems, ions and excited species tend to be formed in groups containing, on the average, 3 ions and about 6 excited molecules. Such groupings, with diameter about 20 A, are called spurs. The spacing between spurs varies from 1–2 molecular diameters for heavy-particle irradiation to thousands of molecular diameters for fast electrons. The existence and distribution of spurs affects the chemistry.

In liquid water, for example, the approximate value of G for products at an early stage of the chemical effects is shown in the accompanying table. In water vapor, the primary yield of molecules decomposed to radicals is estimated as $G \sim 12$.

Radiation	G			
	H	OH	H_2	H_2O_2
Co^{60} γ	3.70	2.92	0.39	0.78
Tritium β	3.06	1.95	0.325	0.88
Polonium α	0.65	0.95	1.50	1.35

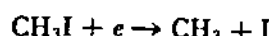
Representative processes. Processes particularly characteristic of radiation chemistry, and some illustrative examples suggested in explanation of various observations include the following (asterisk denotes a molecule in an excited state).

Ion-molecule reactions



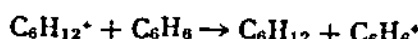
$$CH_3^+ + CH_4 \rightarrow C_2H_5^+ + H_2$$

Dissociative capture of an electron



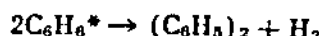
$$CH_3I + e \rightarrow CH_3 + I$$

Charge transfer



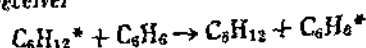
$$C_6H_{12}^+ + C_6H_6 \rightarrow C_6H_{12} + C_6H_6^+$$

Stern-Volmer reactions of excited molecules



$$2C_6H_6^* \rightarrow (C_6H_5)_2 + H_2$$

Protection by energy transfer to a chemically able receiver



$$C_6H_{12}^* + C_6H_6 \rightarrow C_6H_{12} + C_6H_6^*$$

Other nrocesses include some that are observ ble al ation onversion

adicals differ from those of photochemisi photochemistry radicals are formed initially only in pairs In radiation chemistry the existence of spurs results in primary production of four or more free radicals in close proximity See PHOTOCHEM ISTRY

Rad ... *Chemical* effects of high en ergy ... reacte ... ner e ... tions cancer production and so on) Suc ... may be deliberately employed to induce polymeri zation of special kinds to cross link and to ther mall ... polymers to sterilize foods me di ... prot ... pos ... unusual environmental conditions such as ex tremely low temperature *very thick* layers and so on *See* RADIATION BIOLOGY RADIATION DAM AGE (INANIMATE MATERIALS) RADIATION INJURY (BIOLOGY) [M BU]

Bibliography M Burton Introduction to radia tion chemistry *J Chem Educ* 28(8) 404–420 ... nisms of ... *J Phys* ... on W H Hamill and J L Magee A *consideration* of ele mentary processes in radiation chemistry Proc *Second Intern Conf Peaceful Uses Atomic Energy* 29 391–399 1958 J J Harwood et al (eds) *Effects of Radiation on Materials* 1958 G J Hine and G L Brownell (eds) *Radiation Dosimetry* 1956 J L Magee Radiation chemistry *Ann Rev Nuclear Sci* 3 171 192 1953

Radiation cytology

... ive treatment permitting meaningful gen eralizations to be made While the lethal effect of these rad ations on both cell and organism is also a topic of importance any consideration of the other two aspects includes it since lethality is largely a consequence of nuclear and cytoplasmic damage

Materials customarily employed in the chromosomal stud es include the following microspores of *Tradescantia* (spiderwort) microsporocytes of *Trillium* (wakerobin) spermatocytes of grasshoppers root tip cells of the broad bean and onion and the salivary gland cells of *Drosophila* (fruitfly) Mitotic rate studies have been most satisfactorily carried out on the neuroblasts of the grasshopper embryo and the root tip cells of the broad bean but tissue culture techniques now permit adequate ex periments to be performed on mammalian and hu man cells

The induction of chromosomal aberrations and the problem of cell division and mitotic rates have been studied by the means of chromosome breaks induced by radiation

Chromosome breaks Two general types of chromosome breaks induced by radiation are classi fied according to whether the chromosome is longi tudinally single or double at the time of exposure If it is single the break is referred to as chromosomal since the unit of breakage is the chromosome if double the term *chromatid break* is em ployed since the chromatid rather than the whole chromosome is the unit of breakage. A distinction however must be made between a break and an abberation The former implies only a rupture in the chromatin strand while the latter in most in stances involves also the union of broken ends to form recognizable alterations in the structure of the chromosomes such as deficiencies duplications translocations and inversions *See* CHROMOSOME ABERRATION

Relationship to radiation dosage Breaks in chromosomes induced by ionizing radiations are be lieved to be linearly related to dosage and inde pendent of intensity the individual break being caused by the passage of an ionizing particle through or near to the chromatin strand There is no radiation threshold below which they do not ap pear Breaks produced by ultraviolet light show a similar relation to dosage and intensity but the action spectrum indicates that this radiation must be absorbed in the nucleic acids of the chromosome in order to be effective A comparison of the effectiveness of the different types of ionizing ra diations shows that their relative biological effi ciency (RBE) in inducing breaks is closely corre lated with the density of ion pairs The order of effectiveness per unit of dose from least to great est among the commonly used ionizing radiations would be gamma rays hard x rays soft x rays neutrons and alpha rays Thus the greater the mass of the particle and the slower its speed as it moves through the cell the greater its efficiency in break ing chromosomes For example 1 Mev neutrons have an RBE of 2.5 compared with an RBE of 1 0 for hard x rays *See* RADIATION

Oxygen effect The breakage of chromosomes is also related to the amount of oxygen available in the cell at the time of exposure Radiation under anoxic conditions (lack of oxygen) drastically re duces the frequency of breaks The oxygen effect is believed due to diffusible active radicals or mole cules (H OH HO_2 or H_2O_2) formed in the vicin ity of the chromosome when water is decomposed

by radiation, with oxygen playing some as yet undetermined role either in their formation or their action Since the oxygen effect is greatest with gamma rays and x rays, less so with neutrons, and negligibly so with alpha rays, there is an inverse correlation between the role of oxygen in determining breakage and the density of radiation induced ion pairs When the number of ion pairs per micron of path length exceeds 300, the oxygen effect disappears

Chromosome restitution and rejoining Calculations have revealed that 90–95% of the radiation induced breaks are not realized as detectable changes in the chromosome, restitution of broken ends occurs to restore the original alignment of the chromosome When restitution fails, the broken fragment of chromatin is lost, along with the genes it contains, and the chromosome has suffered a deficiency More complex aberrations are formed, however when two or more broken ends, either in the same chromosome or in different chromosomes, rejoin to create new arrangements of chromatin These may be rings, dicentrics or polycentrics possessing two or more centromeres, or acentrics possessing no centromere and consequently, incapable of movement on the spindle during cell division These chromatin rearrangements lead to abnormal anaphase segregations (in mitosis and meiosis), and are largely responsible for cell death Viable rearrangements include inversions and reciprocal translocations *See* MEIOSIS, MITOSIS

Since most complex rearrangements induced by gamma rays or x rays involve two or more independently produced breaks they increase as the square of the dose, unlike deficiencies, which exhibit a linear relationship However restitution, which decreases the number of broken ends in the cell and rejoining which leads to the formation of aberrations by linking broken ends into new arrangements of chromatin are competitive systems Since most broken ends do not remain open indefinitely, and since aberrations involve two independently formed breaks the dose-squared relationship is realized only when the intensity of radiation is high (above 100–150 roentgens per minute) When the intensity is lowered there is a tendency for the frequency of aberrations to decrease

The frequency of aberrations induced by neutrons and alpha rays is determined only by the dosage, no intensity factor is encountered The reason is that the density of ion pair production is sufficiently great to break two or more chromosomes simultaneously, and a linear relation to dosage is therefore maintained

The rejoining of broken ends is an energy requiring event dependent upon the presence of oxygen in the cell, consequently broken ends, when once induced can be kept open by conditions or chemicals which interfere with oxidative metabolism Since anoxia, potassium cyanide (KCN), dinitrophenol (DNP), and carbon dioxide (CO_2) in the dark are effective in this respect both respiration and oxidative phosphorylation are implicated, although the specific links in the chain of reactions remain unclarified Rejoining takes place therefore, only when sufficient energy, probably in the form of high energy phosphate bonds, is available in the cell, in the absence of an energy yielding system, which itself is damaged by radiation, the breaks simply accumulate until normal metabolic conditions permit reactions to proceed

Radiation sensitivity factors The sensitivity of cells to radiation, as reflected in chromosomal damage, is a function of the stage of division the kind of cell, and its metabolic state In general the more actively dividing a tissue is, or the more active a cell is metabolically, the more sensitive it is to radiation damage A deficiency of calcium in a cell also increases its radiation sensitivity Resting cells are relatively insensitive but radiation sensitivity varies during the course of cell division The most sensitive stages appear to be late prophase and metaphase, although the physical reasons determining this variability are undetermined The ratio of sensitivity between metaphase and interphase may be as high as 50 1, with somewhat lower ratios prevailing for other stages

Effect on cell division. Mitotic activity is sharply depressed by both ionizing and photochemical radiations An inhibition of mitosis in grasshopper neuroblasts can be observed after exposure to 4 roentgen (r) of x ray yet 8000 r will not completely suppress cell division Heavy doses of radiation however cause abnormal mitoses, and cell death usually ensues after the completion of the division processes

It has been shown that a depression of cell division by moderate doses of x rays is followed by a compensatory wave of dividing cells bringing the

critical period appears to be late prophase just prior to the breakdown of the nuclear membrane Cells which have passed this stage at the time of exposure are not appreciably inhibited, radiation damage is expressed at the next division Cells in earlier prophase stages eventually exhibit a delayed mitotic activity but an abnormal behavior usually results from chromosomal damage rather than from a malfunction of the mitotic apparatus of the cell A state of anoxia tends to minimize the x ray effects on cell division, leading to the belief that molecular oxygen interacting with the products of decomposed water is instrumental in effecting cell damage

Interference with mitosis can also be brought about by exposure prior to or during the period of chromosomal nucleic acid synthesis in interphase Until the synthetic processes recover the passage of a cell into a state of division is prevented

Ultraviolet radiation can also interfere with cell division but inhibition is not followed by either regression to earlier stages or b a compensatory

wave of dividing cells An action spectrum suggests that inhibition is largely a cytoplasmic phenomenon since radiation effect of 2240 angstroms (A) is more deleterious to cell division than is the effect of 2537 A Therefore the effect of ultraviolet is on the proteins of the cytoplasm rather than on the r cleic acids of the nucleus and it seems likely ... apparatus ... RA ... SW]

Radiation damage (inanimate materials)

Harmful changes in the properties of liquids, gases, and solids caused by interaction with nuclear radiations Interest in radiation damage to inanimate materials is almost entirely limited to materials that are used structurally or otherwise within the radiation field of a nuclear reactor For a description of damage caused to biological systems by radiation, see RADIATION INJURY (BIOLOGY)

Radiation damage in nuclear reactors is caused by high energy radiation created by the fissioning of uranium 235 The most important agents in causing damage are γ rays β rays, highly energetic (fast) neutrons and fission fragments The average energy of each of these types of radiation is approximately 1 000 000 electron volts (1 Mev) except for fission fragments where the energy shared between the two fragments is approximately 160 Mev *See* BETA RAYS FISSION NUCLEAR GAMMA RAYS NEUTRON

While all details of radiation damage are not understood it has often been found possible to minimize or to eliminate radiation damage in reactors by choice of material increase in temperature of irradiation proper treatment prior to irradiation and improved heat transfer

Damage mechanisms There are several ways in which radiation can interact with matter to cause damage

All the radiation mentioned with the exception of neutrons can directly cause ionization and electronic excitation These two effects can create chemical changes or increased chemical reactivity or both in most materials other than already ionized materials or metals.

Effects of neutrons The neutron cannot directly create ionization or excitation because of its neutral character However a fast neutron can cause damage by interacting in a 'billiard ball' collision with a nucleus The nucleus then recoils, some of the neutron energy having been transferred to it In the case of an un ionized gas liquid or solid this recoil nucleus which is a charged ion can then cause excitation and ionization The covalent hydrocarbons are particularly subject to damage from recoil nuclei In the case of an ionized crystalline solid the recoil nucleus is displaced from the crystal lattice and can in turn displace additional atoms from their normal positions Thus a cascade of excited ionized or displaced atoms can be set up

One other effect of neutrons that can sometimes be of importance is the transmutation of the struck atom to form a new chemical species This effect arises from the ability of many atomic nuclei to absorb low energy (thermal) neutrons, which have an energy of 0 025 ev for the most probable neutron velo ... °C The probability of thermal neutron ... to as its ... (see NEUTRON CROSS ... the newly formed nucleus is in an excited, radioactive state When the thermal neutron absorption cross section is large, an appreciable quantity of impurity may be introduced by neutron capture after prolonged irradiation in a nuclear reactor This induced radioactivity, often long lived, means that postirradiation studies must be done remotely in heavily shielded enclosures.

The β rays, γ-rays, and fission fragments can also cause the displacement of large numbers of atoms, though the number displaced by β and

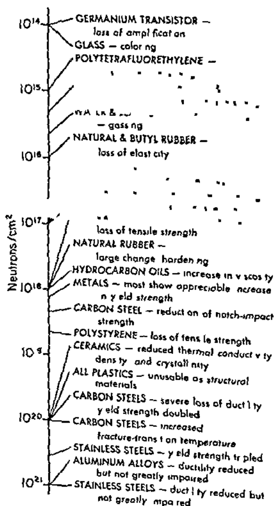


Fig 1 Sensitivity of engineering materials to radiation Levels indicated are approximate and subject to variation Indicated changes are in most cases at least 10% (After O Sisman and J C. Wilson, Nucleonics 14 58–65, 1956)

γ rays is small compared to those displaced by fast neutrons and fission fragments.

Extent of damage Electronic excitation and ionization is the most important damage mechanism in gases and liquids, while the displaced atom mechanism is most important in solids

There are several important factors that determine the extent of radiation damage. (1) the energy and intensity of the particles or radiation, (2) the temperature at which irradiation takes place, and (3) the type of material being irradiated

Figure 1 shows the relative susceptibility of various types of material to radiation damage When it is considered that several reactors now operate at a flux of 10^{14} neutron/(cm²)(sec), it can be seen that the useful lifetime of many of the materials listed will be extremely short Metals in general are most resistant to damage, but this resistance is balanced by the fact that, as structural materials of the reactor core, they are usually subjected to the most intense irradiation Water and organic liquids are the most sensitive, since they are subject to decomposition or change by the formation of free radicals, chain rupture, polymerization, and depolymerization

Damage to solids. The average energy required to displace an atom from its normal position is 25–35 ev Since the fission neutron has an average energy of 10^6 ev, the energy imparted to the displaced atom may be many times the minimum energy required for displacement The amount of energy transferred to the displaced atom is a function of the angle of impact of the incoming particle, its energy, and the mass of the struck-on atom Thus, a hydrogen atom could absorb all the neutron's energy in a head on collision In general, the average energy transferred is

$$\Delta E = \frac{2Mm}{(M+m)^2} \quad E$$

where M = mass of knocked on atom, m = mass of neutron, and E = energy of neutron If M is expressed as the mass number or atomic weight of the element, m is approximately unity and the equation reduces numerically to

$$\Delta E = \frac{2M}{(M+1)^2} \quad E$$

The removal of an atom from its equilibrium position and the final lodging of the atom in a non

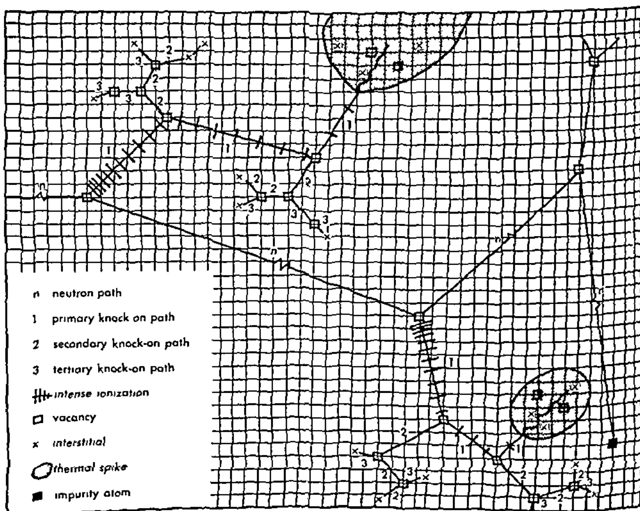


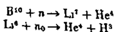
Fig 2 The five principal mechanisms of radiation damage are ionization, vacancies, interstitials impurity atoms, and thermal spikes The diagram shows how a neutron might give rise to each of them in copper Grid line intersections are equilibrium positions for atoms (After D S Billington, Nucleonics 14 54–57, 1956)

equilibrium position can lead to serious changes in the properties of the material since these properties particularly in solids *are a critical* function of the relative position of the atoms to one another in the crystal lattice Thus the strains set up by the atoms lodged in *nonequilibrium* positions (interstitials) may lead to hardening and even embrittlement The vacant lattice positions may, *under appropriate* temperature conditions, lead to the acceleration of unwanted solid state reactions in alloys and other polycomponent solids since the additional vacant lattice sites lead to *more ready* intermingling of the different types of atoms that make up the alloy Figure 2 is a schematic representation of the various mechanisms of damage that take place in a solid

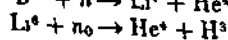
Susceptibility to damage Some components of a nuclear reactor are more subject to damage than others The nuclear fuel itself is the most damaged portion since it is the only part of the reactor that comes in contact with the fission fragments These highly charged high energy massive particles (≈80 Mev each) deposit all their excess energy in a distance of 5–15 μ This leads to thousands of displaced atoms and numerous momentarily high temperature regions within the crystal These hot spots may lead to the equivalent of local melting in regions involving a few thousand atoms at a time The damage in uranium displays itself as swelling embrittlement elongation and weakening of the uranium The fast neutrons and γ rays also contribute to damage of the fuel but these contributions are relatively unimportant *See* NUCLEAR FUELS

Neutrons are able to travel large distances between collisions (on the order of centimeters) and so can penetrate to all parts of the reactor Hence moderator coolant shielding control instrumentation and structural components are all subject to damage by fast neutrons The γ rays also have a small probability of being absorbed in short dis

steels, dissociation of water and organic liquids swelling and reduction in thermal conductivity of ceramics, partial loss of ductility and accelerated decomposition of metastable alloys which in turn may lead to poor corrosion resistance and reduction in permeability of magnetic metals and alloys are typical of the damage that has been observed Control rods when they contain boron 10 or lithium-6 are subject to damage from the recoil products of the reactions



$$B^{10} + n \rightarrow Li^7 + He^4$$



and $$Li^6 + n_0 \rightarrow He^4 + H^3$$

which behave in a similar manner to the fission fragments of uranium 235

Wigner energy One important manifestation of radiation damage is the stored energy that builds up within certain types of materials upon prolonged irradiation This stored energy often called Wigner energy, can be a source of damage if released rapidly For example, graphite, after irradiation has reportedly contained up to 400 cal of stored energy per gram The instantaneous release

ditions or the use of increased temperature of operation have been effective, in the usual case, in a ... of hazard For example, the

high temperature operation constitutes a ... ous self annealing process

Damage to liquids Water is used as a heat transfer medium in many reactors Organic liquids have also been attracting interest as coolants, in addition to their usual role as lubricants In both cases, r... tance to radiation damage is important

a small amou...

taining water purity by means of an ion exchange system

The most striking effects of radiation damage to organic liquids are gas evolution because of decomposition, and increased viscosity because of polymerization Aromatic hydrocarbons, by virtue of their structures, are much more resistant to radiation than the straight chain aliphatic hydrocarbons [D.S.B]

Bibliography G J Dienes and G H Vineyard, *Radiation Effects in Solids* 1957, S. Glasstone, *Principles of Nuclear Reactor Engineering*, 1955, J J Harwood *Effects of Radiation on Materials*, 1958, F Seitz and D Turnbull (eds.), *Solid State Physics* vol 2, 1956

Radiation injury (biology)

The basic building block of life is the cell Hence it is logical to attribute the harmful effects of ionizing radiation on plants and animals to changes in

ISTRY, RADIATION BIOLOGY RADIATION CYTOLOGY

In multicellular organisms effects on cells are complicated by the interaction of injured and intact cells Consequently proper understanding of radiation injury in such organisms calls for appreciation not only of the reaction of individual cells but also of groups of cells in various organs and tissues

EFFECTS OF RADIATION ON CELLS

Effects on genes The most important action of radiation on the cell is the production of changes in the genetic or chromosomal apparatus. Once established these changes are largely irreversible and because of the strategic importance of each gene may alter the fate of the cell and of its prog

eny as well The changes may come about through (1) mutation, or alteration of the gene, which is believed to consist of a change in the nucleoprotein that constitutes the gene When established, it may be reversed or otherwise altered only by further mutation A mutation induced by irradiation is essentially indistinguishable from a natural one Analysis of the relation between mutation rate and radiation dose has led to the conclusion that mutation may require but a single ionization of the gene, the frequency of mutations characteristically rises in linear proportion to the dose without evidence of a threshold (2) Chromosome breakage is visible microscopically as a break in the continuity of the chromosomal thread The ends of the broken structure of the chromosome often reunite within a few minutes, restituting the original structure of the chromosome, but they may not rejoin at all In this latter case, the distal chromosomal fragment and its genes may be lost or the broken ends may unite with fragments of other chromosomes and cause various genetic rearrangements (3) Chromosome stickiness and clumping which presumably affect the chromosomal surface, cause chromosomes to adhere to one another and clump together They thus fail to separate normally at cell division *See* MUTATION

Effects on cell division. Relatively small amounts of radiation postpone cell division, the delay varying with the dose When division is eventually resumed in a population of affected cells the percentage dividing may temporarily exceed or "overshoot" normal, this is often associated with extensive degeneration of the dividing cells If large enough doses are administered the cells' ability to divide is permanently abolished *See* MITOSIS

Killing of cells. Cells differ markedly in their susceptibility to radiation induced death but any cell may be killed if it receives enough radiation In general susceptibility varies in proportion to the rate of cell division few rapidly dividing mammalian cells being able to survive 500 rad of x rays (A rad is a unit of absorbed radiation dose, one rad equaling 100 ergs per gram of absorbing tissue) Although radiation death may occur during or immediately after irradiation more often it ensues when the cell attempts to divide or after it has divided several times The relation between the percentage of cells surviving and the radiation dose implies that the dose required to kill all cells in a given population varies with the total number of cells in that population Hence although the median lethal dose per cell may be less than 100 rad several thousand rad may be required to kill all the billions of cells in a large organ or tumor This is attributable to the spatial distribution of radiation induced ion pairs in the absorbing medium the probability of all cells being appropriately affected decreasing with the total number of cells irradiated

The mechanism by which radiation kills cells is unknown, but the high radiosensitivity of dividing cells, with their tendency to die during cell division, suggests that the most critical type of injury involves the reproductive or chromosomal apparatus of the cell This is also suggested by the fact that many irradiated cells rendered incapable of division retain their ability to differentiate into more highly specialized forms and to synthesize nucleic acid and protein Enormous amounts of radiation (many thousand rad) are often required to stop metabolic activity in such nondividing cells

EFFECTS OF RADIATION ON ANIMAL TISSUES

Tissues differ widely in radiosensitivity but in general their susceptibility to radiation varies with the division rate of their component cells Accordingly, the most radiosensitive tissues of the body composed of cells that divide rapidly are the blood forming organs, gonads, skin and intestine

Radiation injury is not however limited to the killing of radiosensitive cells Invariably, the initial cellular destruction leads to secondary disturbances and reparative processes, often through systemic mechanisms, which modify the primary lesion In this respect, radiation injury simulates other types of injury and is not unique or specific.

Blood forming tissues. Since circulating blood cells live only a few days or weeks, they must be replaced continually The new cells are produced in the bone marrow, spleen, and lymphoid organs through division of blood cell precursors which are highly radiosensitive and hence readily destroyed by radiation Because destruction of these cells leads to shortage of mature blood cells, which may have drastic consequences, damage to the blood forming organs is one of the most important types of radiation injury *See* HEMATOPOIESIS

If only a part of the body is irradiated blood forming cells from nonirradiated tissue are carried by the blood stream to the damaged organs where they multiply and replace destroyed cells. Consequently greater quantities of radiation are toler

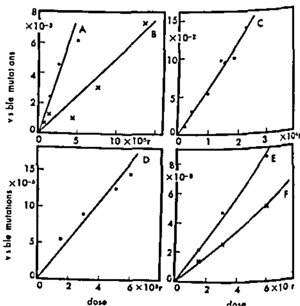


Fig 1 Proportion of visible mutations induced by x rays in relation to dose A B tobacco mosaic virus C, *Neurospora* D F, *Drosophila melanogaster* (*From D E Lea, Actions of Radiations on Living Cells* 2d ed, Cambridge, 1955)

ated if applied to part of the body than if applied

pecies, and the radiation dose in general ... er of cells is depleted more rapidly in small lab ratory animals than in man, the depletion being accelerated in any given species with increasing dose of radiation The latter may be explained by the fact that blood cells intermediate in matura tion are intermediate in radiosensitivity In the fol lowing remarks the blood changes are character ized according to their time sequence in man

The lymphocyte is the first blood cell to disap pear Diminution in the number of these cells is ... evident within only a few hours after irra

lymphocytes to appear in the blood The loss of lymphocytes is believed to be responsible to some extent for the depression of immunity that occurs soon after intensive irradiation and that predis poses the affected individual to infection

The granulocyte another type of white blood cell disappears more slowly from the blood stream, al though paradoxically the number of such cells in the blood may be increased for several days after irradiation owing to a constitutional reaction of the body to stress Since mature granulocytes are nondividing cells and are relatively radioresistant, the gradual depletion of their number reflects pre dominantly the loss of aging cells from the total population the maximal depression of the granulo cyte count not being reached until 5 6 weeks after exposure Because these cells constitute the body s most important defense against bacterial invasion infection is a common cause of death in heavily ir radiated individuals

Simultaneously with the fall in granulocyte count, the number of circulating blood platelets also declines, maximal deficiency of these corpus cles occurring 3-5 *weeks after irradiation* Depres sion of the platelet count below a certain level leads to hemorrhage The bleeding which may be either localized or generalized, varies in severity and can be fatal

The red blood cell count falls gradually in the absence of hemorrhage severe anemia occurring 5 or more weeks after irradiation The anemia re sults primarily from underproduction of new red blood cells owing to destruction of precursors in the bone marrow Although red blood cell forming elements are usually among the first to regenerate after radiation injury regeneration may be delayed, and the anemia may therefore be severe and pro longed even in the absence of bleeding.

Skin The response of the skin to radiation has been studied extensively because the skin is the most exposed structure of the body and because its radiosensitivity has until recently seriously limited

the ...

from transient reddening a reaction ... to measure dose in radiotherapy, to ulceration and ... is characteristi

of the underlying connective tissue, and ... cutaneous blood vessels may gradually appear These changes, known collectively as *radiation der matitis* predispose to subsequent development of skin cancer

Injury of accessory skin structures is manifested in graying and loss of the hair The epilation or loss of hair, appears several weeks after irradia tion and may be temporary or permanent depend

to the abdomen elicit profound effects These vary from slight disturbances of motility and secretion to ulceration and sloughing of the lining of the bowel

Nausea and vomiting usually ensue within a few hours after exposure to doses above 200 rad After several times this much radiation, sloughing of the intestinal lining may lead to ulceration, intractable diarrhea dehydration and invasion of the blood stream by bacteria that normally inhabit the lumen of the bowel This sequence of events is usually fatal and constitutes one of the major causes of death after massive irradiation of the whole body

diately however, owing to survival of more-radio-resistant mature eggs or sperm but only after these preformed cells are eliminated from the body Even then sterility is transitory unless too few precur sors survive to resume adequate production of germ cells In man, as in most other mammals studied permanent sterilization requires amounts of radia tion that are lethal when absorbed by the entire body hence sterility is not generally a complica tion of whole body irradiation

Apart from killing the germ cells mutations may be induced by radiation and be passed on via the eggs and sperm to successive generations of prog eny These genetic disturbances are thought to re sult from the minutest amounts of radiation and are therefore considered to be the limiting factor in the radiation dose permissible for mankind They are discussed further in this article under genetic effects of radiation

Eye. The part of the eye most easily injured by radiation is the lens opacification of which (cata ract formation) has been observed after only 200 rad of x rays Even smaller doses of neutrons are estimated to have caused cataracts, several exam ples of such induction have been noted among neu

tron physicists The cornea conjunctiva, and the retina withstand much more radiation The retina however is highly radiosensitive early in its embryonic development Through radiochemical reactions in the retina which are harmless minute amounts of ionizing radiation are visible *See* EYE

Nervous system. Although only relatively large amounts of radiation will kill nerve cells in the adult, the developing nervous system is highly radiosensitive Even in the adult, transitory functional disturbances may be elicited by relatively low doses, and after intensive exposure of the brain (1000–10 000 rad depending on the species), incapacitating neurological effects may lead to death within minutes or hours *See* NERVOUS SYSTEM

Bones and teeth Only a few hundred rad applied to bone and tooth forming cells in infancy or early childhood cause disturbances of dentition and skeletal growth In contrast, mature bones and teeth are relatively radioresistant Large amounts of radiation however such as may accumulate from locally deposited radioisotopes or from the treatment of cancer produce demineralization and necrosis of bone that lead to fractures, loosening of the teeth, bone cancer, and other complications *See* BONE

Vascular system Transitory dilation of blood vessels causing erythema or reddening of the skin is one of the earliest known reactions to ionizing radiation It occurs after only a few hundred rad and may be accompanied by increased permeability of blood capillaries Larger amounts of radiation may severely injure or kill cells that line the walls of blood and lymph vessels giving rise to rupture occlusion permanent dilation or scarring of the affected vessels Adverse effects of these changes on the blood supply may lead to secondary effects such as metabolic disturbances and atrophy in involved tissues *See* CARDIOVASCULAR SYSTEM, LYMPHATIC SYSTEM

Endocrine glands. The glands of internal secretion have traditionally been regarded as radioresistant because of their ability to withstand relatively large amounts of radiation without developing morphological lesions There is growing evidence however, that rather small doses may elicit changes in endocrine function and may, in some instances, induce lasting functional impairment Apart from radiation injury itself, the endocrine system s adaptational response to acute effects of radiation resembles its response to other types of stress. *See* ENDOCRINE SYSTEM

Urinary system. The kidney and lower urinary tract are relatively radioresistant. Depending on the species, however, doses in excess of 500-2000 rad may cause gradually progressive scarring and atrophy of the kidney, which lead to fatal loss of renal function

Lungs. Although relatively radioresistant, the lungs may be injured by intensive irradiation The resulting injury consists of a chronic, pneumonia like disease with scarring of lung tissue and blood vessels A complication from locally deposited radioisotopes is cancer of the lung which has been noted in miners of radioactive ore and in experimental animals

EFFECTS OF WHOLE BODY IRRADIATION

Acute radiation syndrome When the entire body is irradiated killing of cells in the various radiosensitive organs causes a complex series of disturbances the predominant signs and symptoms of which are referable to injury of the intestinal tract blood forming organs and skin The injury if severe enough is fatal within 30 days in most laboratory animals but in man death may be delayed until the second month after irradiation has occurred

Susceptibility to death varies from species to species the average median lethal radiation dose for mammals being about 500 rad As with most other lethal agents radiation in doses below a certain minimum threshold causes negligible mortality but in doses above a maximum level is uniformly lethal (Fig 2)

The cause of death varies with the species dose, dose rate and type of radiation The predominant cause of death in most mammals is injury of blood forming tissues which results in infection hemorrhage and anemia Lethal injury of the intestine

Table 1 Symptoms of acute radiation syndrome*

Time after exposure	Supralethal dose 1000 rad	Median lethal dose 400 rad	Sublethal dose 200 rad
First week	Nausea and vomiting first day	Nausea and vomiting first day	
Second week	Continued nausea vomiting diarrhea fever inflammation of throat prostration emaciation leading to death		
Third week		General malaise loss of appetite loss of hair hemorrhage pallor diarrhea fever inflammation of throat emaciation leading to death in 50% of victims	Loss of appetite loss of hair inflammation of throat pallor hemorrhage diarrhea recovery begins no deaths in absence of complications

* Modified from P Alexander *Atomic Radiation and Life* Penguin 1957

Table 2. Median lethal dose of x rays for various species of organisms*

Organism	Median lethal dose, rad
Viruses	
Tobacco mosaic	200 000
Rabbit papilloma	100 000
Bacteria	
E coli	5 000
B mesentericus	130 000
Algae	
Mesolenium	8 500
Pandorina	4 000
Protozoa	
Colpidium	330 000
Paramecium	300 000
Vertebrates	
Goldfish	750
Mouse	450
Rabbit	800
Rat	600
Monkey	450
Man (?)	400

* Modified from E. Paterson, in R. Paterson (ed.), *The Treatment of Malignant Disease by Radium and X Rays*, E. Arnold, 1948.

usually requiring somewhat higher doses than lethal injury of the marrow, is another major cause of death. In this case death results from diarrhea, dehydration, loss of body salts, and massive bacterial invasion from the lumen of the bowel. With even larger doses of radiation, death may ensue rapidly from injury of the brain, as mentioned ear-

l

injury of any one of a variety of organs is potentially lethal.

Shortening of the life span. Animals surviving the acute radiation syndrome usually appear outwardly to have recovered and to be normal by the second or third month after irradiation. That such ... lly recovered, however, is indi-

life span ... irradiation

The bas... tion ... clear, but the ev... gests ... deger ... sociate...

served repeatedly in irradiated rodents, there is not yet conclusive evidence that life-shortening occurs in man. Other late effects, however, have been noted in man, for example, the development of radiation nephritis, aplastic anemia, leukemia, and other types of cancer.

RADIATION INDUCED CANCER

Historical. It is paradoxical that ionizing radiation, which is a potent weapon in the treatment of cancer, should also be capable of causing cancer. As such, however, it is but one of many agents, including ultraviolet rays, viruses, and a variety of chemicals, that are known to have cancer inducing potency. The earliest known example of radiation induced cancer was reported in 1902, less than 10 years after the discovery of x rays. Since then, numerous instances have been observed in man, and the process has been studied extensively in experimental animals.

Although susceptibility to cancer induction varies widely among species and organs, virtually all

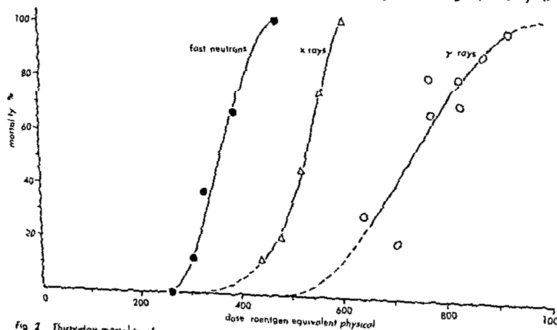


Fig. 2. Thirty-day mortality of mice exposed to x rays, fast neutrons, and γ-rays. The mice were 9–12 weeks old at the time of whole body irradiation. (From A. C. Upton et al., Radiation Research, vol. 4, Academic Press, 1956)

Table 3 Incidence of leukemia in Japanese atomic bomb survivors in Hiroshima

Distance from hypocenter, meters	Average max dose of radiation, rem*	No of survivors†	Survivors with leukemia‡ No	Survivors with leukemia‡ %
0–999	1300	1 200	18	1 500
1000–1499	500	10 500	39	0 371
1500–1999	50	18,700	9	0 046
2000 and over		67,700	12	0 017

* Crude approximation of radiation dose according to method of E B Lewis (1957) These figures, because of their great uncertainty can be considered to provide at best only a semiquantitative correlation between leukemia incidence and dose

† From National Academy of Sciences, *Pathologic Effects of Atomic Radiation* Natl Research Council Publ No 452, 1956

‡ From N Wald, Leukemia in Hiroshima city atomic bomb survivors, *Science*, 127 699–700, 1958

types of cancer have been induced experimentally, and it is clear that all types of ionizing radiation share cancer forming potency The radiation induced cancers that have been noted most often in man are cancer of the skin, leukemia, cancer of bone, and cancer of the lung

Cancer of the skin was encountered early in this century as a complication of radiation dermatitis, developing on the hands and fingers of many of the pioneer radiologists This effect resulted from prolonged manipulation of radiation equipment, the hands being exposed to large cumulative doses of radiation in the era before the attendant hazards were suspected Although this disease claimed the lives of more than 100 of the first workers in radiology, because of present safeguards it is no longer an occupational threat

Another malignant disease encountered with unusual frequency in radiologists is leukemia This disease is also abnormally prevalent in other populations exposed to radiation, such as the Japanese atomic bomb survivors

Radiation induced bone cancer was first noted in painters of luminous watch dials, who inadvertently ingested toxic quantities of radium containing paint by habitually drawing their paint brushes to a point between their lips Deposition of the radium in the skeleton where it delivered relatively large doses of radiation to small foci of bone led to the development of skeletal cancer Osseous tumors are also abnormally prevalent in those who have consumed radium for medicinal purposes

Cancer of the lung in miners employed in the pitchblende mines of Saxony, traditionally the predominant cause of death among these workers has been attributed to their prolonged exposure to radon which is present in high concentrations in the air of such mines (7×10^{-12} to 7×10^{-9} curie per liter) It is noteworthy, however, that many other complicating factors present in this population may have contributed to the effects of radiation in causing lung cancer

Cancer incidence versus radiation dose. Although radiation increases the incidence of many types of cancer, the precise relation between incidence and dose is not known for any type of cancer, especially at low radiation levels The yield of tumors per unit dose cannot, therefore, be accurately estimated, nor can it be assumed that any amount of radiation, however small, will increase the cancer rate Neither can it be assumed that a threshold does exist below which no cancers are induced Since, however, for most experimentally induced cancers the incidence is not linearly proportional to dose, the evidence suggests that at very low dose levels radiation is appreciably less effective than at high dose levels, if it is effective at all Paradoxically, irradiation actually reduces the frequency of certain types of tumors in experimental animals Accordingly, no generalization about the relation between cancer incidence and radiation dose can be applied to all types of cancer

Effect of constitutional variables. Susceptibility to cancer induction in experimental animals varies widely with constitutional variables such as genetic inheritance, age at irradiation, sex, and hormonal activity Furthermore, the influence of any one of these variables on susceptibility to tumor formation often differs from the influence of others and from one type of tumor to another As yet, little is known about the action of these variables on susceptibility in experimental animals and almost nothing about their action in man

Mechanisms of cancer induction. The essential change caused by radiation that leads ultimately to the development of cancer is unknown, as is the

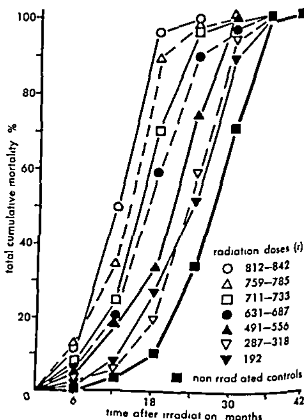
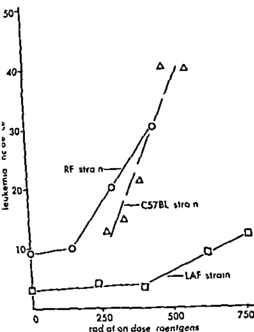


Fig 3 Longevity of female mice surviving 30 days after exposure to atomic bomb γ-rays at 6–12 weeks of age (From A C Upton, J Gerontol, 12 307–313, 1957)



F g 4. Leukem a incidence in mice of several inbred stra ns in relation to dose The data include only those leukemias arising in the thymus as lymphosarcomas. In all instances the m ce were exposed to a single dose of whole-body rad ation early in adult l fe

nature of the cancer change itself There is grow ing evidence however that the transformation of a normal cell into a cancer cell entails a series of alterations and is not merely a one-step process This is suggested by the relatively long induction period intervening between irradiation and the on set of malignant growth a period averaging 5–20 years in man depending on the type of cancer in duced It is also suggested by the tendency of nat urally occurring cancer to develop late in life and to become progressively more frequent with advanc ing age

One theory of cancer formation ascribes the can cer change to one or more somatic mutations in the

ized to the site of tumor formation such as skin or bone in other situations radiation appears to initiate the development of cancer in nonirradiated cells by indirect mechanisms that is through ex cessive stimulation of their growth for reparative purposes In the latter case although cancer may

ated Instead the cancer seems to arise through se lection of spontaneously occurring cancer mutants by a radiation induced change favoring the growth of such mutants

Another theory ascribes the cancer forming ef fects of radiation to activation of cancer viruses This hypothesis depicts radiation as (1) increas ing susceptibility to cancer viruses by depression of immunity, (2) unmasking or activating latent cancer viruses through mechanisms yet to be dis closed, or (3) causing formation of new viruses with cancer forming potency through interference with nucleoprotein synthesis in the normal cell As yet, viruses have not been definitely implicated in the development of any radiation induced tumors, but there is conclusive evidence that they give rise to many types of cancer in nonirradiated animals of diverse species

EFFECTS ON THE EMBRYO AND FETUS

Although radiation injury of human embryos is documented most of the knowledge of the effects of radiation on embryonic and fetal development comes from experiments with laboratory animals These studies have shown that the individual is more vulnerable to killing by radiation early in development than at any other time of life. The most sensitive period is that preceding implanta tion of the fertilized egg in the uterus. At this stage death may be caused by less than one third ... to kill the adult animal After

schedule Malformation of almost any organ may be induced by irradiation at the appropriate moment during or preceding its development Since the ma jor period of organ formation occurs early in em bryonic life (during the first trimester of pregnancy in humans), susceptibility to gross malformations decreases in the middle and latter part of gesta tion Even late in gestation however, irradiation may cause abnormalities in the development of slowly maturing organs such as the brain, eye, and gonads Subtle degrees of these abnormalities may not be evident at birth

How radiation causes malformations is not yet fully known although interference with organ de velopment through the killing of embryonal cells must play an important role in the process What ever the mechanism of malformation the sensitive period for induction of a given abnormality is fre quently very short With increasing radiation dose, however the sensitive period tends to be prolonged and the incidence and severity of abnormalities increased Doses as low as 25 rad have been shown experimentally to cause developmental dis turbances if applied at the critical time For this reason, care should be taken to avoid exposing the abdomen to radiation during early pregnancy

In addition to malformations other delayed in juries of the type induced by irradiation in adult life may be produced by exposure in utero These

Table 3 Incidence of leukemia in Japanese atomic bomb survivors in Hiroshima

Distance from hypocenter, meters	Average max dose of radiation, rem*	No of survivors†	Survivors with leukemia‡ No	Survivors with leukemia‡ %
0-999	1300	1,200	18	1 500
1000-1499	500	10 500	39	0 371
1500-1999	50	18 700	9	0 046
2000 and over		67,700	12	0 017

* Crude approximation of radiation dose according to method of E B Lewis (1957) These figures, because of their great uncertainty, can be considered to provide at best only a semiquantitative correlation between leukemia incidence and dose

† From National Academy of Sciences, *Pathologic Effects of Atomic Radiation*, Natl Research Council Publ No 452, 1956

‡ From N Wald Leukemia in Hiroshima city atomic bomb survivors, *Science*, 127 699–700, 1958

types of cancer have been induced experimentally, and it is clear that all types of ionizing radiation share cancer forming potency The radiation induced cancers that have been noted most often in man are cancer of the skin, leukemia, cancer of bone, and cancer of the lung

Cancer of the skin was encountered early in this century as a complication of radiation dermatitis, developing on the hands and fingers of many of the pioneer radiologists This effect resulted from prolonged manipulation of radiation equipment, the hands being exposed to large cumulative doses of radiation in the era before the attendant hazards were suspected Although this disease claimed the lives of more than 100 of the first workers in radiology, because of present safeguards it is no longer an occupational threat

Another malignant disease encountered with unusual frequency in radiologists is leukemia This disease is also abnormally prevalent in other populations exposed to radiation, such as the Japanese atomic bomb survivors

Radiation induced bone cancer was first noted in painters of luminous watch dials, who inadvertently ingested toxic quantities of radium containing paint by habitually drawing their paint brushes to a point between their lips Deposition of the radium in the skeleton where it delivered relatively large doses of radiation to small foci of bone led to the development of skeletal cancer Osseous tumors are also abnormally prevalent in those who have consumed radium for medicinal purposes

Cancer of the lung in miners employed in the pitchblende mines of Saxony, traditionally the predominant cause of death among these workers has been attributed to their prolonged exposure to radon, which is present in high concentrations in the air of such mines (7×10^{12} to 7×10^{9} curie per liter) It is noteworthy however that many other complicating factors present in this population may have contributed to the effects of radiation in causing lung cancer

Cancer incidence versus radiation dose. Although radiation increases the incidence of many types of cancer, the precise relation between incidence and dose is not known for any type of cancer, especially at low radiation levels. The yield of tumors per unit dose cannot, therefore, be accurately estimated, nor can it be assumed that any amount of radiation, however small, will increase the cancer rate Neither can it be assumed that a threshold does exist below which no cancers are induced Since, however, for most experimentally induced cancers the incidence is not linearly proportional to dose, the evidence suggests that at very low dose levels radiation is appreciably less effective than at high dose levels, if it is effective at all Paradoxically, irradiation actually reduces the frequency of certain types of tumors in experimental animals Accordingly, no generalization about the relation between cancer incidence and radiation dose can be applied to all types of cancer

Effect of constitutional variables. Susceptibility to cancer induction in experimental animals varies widely with constitutional variables such as genetic inheritance, age at irradiation, sex, and hormonal activity Furthermore, the influence of any one of these variables on susceptibility to tumor formation often differs from the influence of others and from one type of tumor to another As yet, little is known about the action of these variables on susceptibility in experimental animals and almost nothing about their action in man

Mechanisms of cancer induction. The essential change caused by radiation that leads ultimately to the development of cancer is unknown, as is the

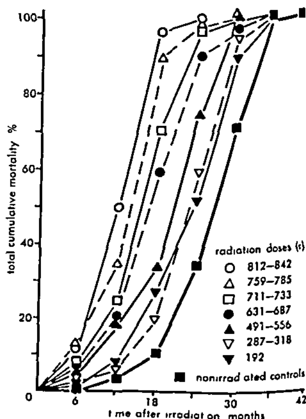


Fig 3 Longevity of female mice surviving 30 days after exposure to atomic bomb γ-rays at 6–12 weeks of age (From A C Upton, *J Gerontol*, 12 307–313, 1957)

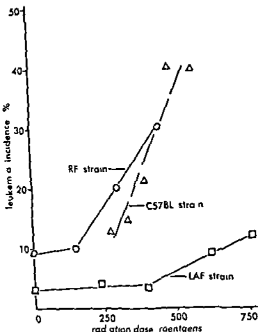


Fig 4 Leukemia incidence in mice of several inbred strains in relation to dose The data include only those *leukemias arising in the thymus as lymphosarcomas.* In all instances the mice were exposed to a single dose of whole body radiation early in adult life

nature of the cancer change itself There is growing evidence however, that the transformation of a normal cell into a cancer cell entails a series of ... step process

set of malignant growth, a period averaging 5–20 years in man depending on the type of cancer induced It is also suggested by the tendency of naturally *occurring cancer to develop late in life* and to become progressively more frequent with advancing age

One theory of cancer formation ascribes the cancer change to *one or more somatic mutations* in the cell which predispose it to unrestrained growth According to this view the cancer forming action of ionizing radiation stems from its mutagenic potency Although this hypothesis *may apply in instances* in which cancer is induced by effects localized to the site of tumor formation such as skin or bone in other situations radiation appears to initiate the development of cancer *in nonirradiated* cells by indirect mechanisms that is, through excessive stimulation of their growth for reparative purposes *In* the latter case, although cancer may conceivably arise through somatic *mutations* these are not induced by the mutagenic action of radiation since the cancer forming cells are not irradiated Instead, the *cancer* seems to arise through selection of spontaneously occurring cancer *mutants* by a radiation induced change favoring the growth of such mutants

Another *theory ascribes* the cancer forming effects of radiation to activation of cancer viruses. This hypothesis depicts radiation as (1) increasing susceptibility to cancer viruses by *depression* of immunity, (2) *unmasking or* activating latent cancer *viruses through* mechanisms yet to be disclosed, or (3) causing formation of new viruses with cancer forming potency through interference with nucleoprotein synthesis in the normal *cell* As yet, viruses *have not been definitely* implicated in the *development* of any radiation induced tumors, but there is conclusive evidence that they give rise to many types of cancer in nonirradiated animals of diverse species

EFFECTS ON THE EMBRYO AND FETUS

Although radiation injury of human embryos is documented, most of the knowledge of the effects of radiation on *embryonic and fetal development comes from* experiments with laboratory animals These studies have shown that the individual is more vulnerable to killing by radiation early in development than at any other time of life. The *most sensitive period is that preceding implantation* of the fertilized egg in the uterus. At this stage death may be caused by less than one third the dose required to kill the adult animal After implantation and with increasing *maturation*, susceptibility to killing declines, and death tends to be postponed until birth or thereafter

sequentially according to a definite order and time schedule Malformation of almost any organ may be induced by irradiation at the appropriate moment *during or preceding its* development Since the major period of organ formation occurs *early in embryonic* life (during the first trimester of pregnancy in humans) susceptibility to gross malformations decreases in the middle and latter part of gestation Even late in gestation, *however, irradiation* may cause abnormalities in the development of slowly maturing organs such as the brain, eye, and gonads Subtle degrees of these abnormalities may not be evident at birth

How radiation causes malformations is not yet fully known although interference with organ development through the killing of embryonal cells must play *an important role in* the process. Whatever the mechanism of malformation, the *sensitive* period for induction of a given abnormality is frequently very short With increasing radiation dose, however, the *sensitive period tends* to be prolonged and the incidence and severity of *abnormalities* increased Doses as low as 25 rad have been shown experimentally to cause developmental disturbances if applied at the critical time For this reason, care should be taken to avoid *exposing* the abdomen to radiation during early pregnancy

In addition to malformations, other delayed injuries of the type *induced* by irradiation in adult life may be produced by exposure *in utero* These,

Table 3 Incidence of leukemia in Japanese atomic bomb survivors in Hiroshima

Distance from hypocenter, meters	Average max dose of radiation, rem*	No of survivors†	Survivors with leukemia‡	
			No	%
0–999	1300	1,200	18	1 500
1000–1499	500	10 500	39	0 371
1500–1999	50	18,700	9	0 046
2000 and over		67 700	12	0 017

* Crude approximation of radiation dose according to method of E B Lewis (1957) These figures, because of their great uncertainty, can be considered to provide at best only a semiquantitative correlation between leukemia incidence and dose

† From National Academy of Sciences, *Pathologic Effects of Atomic Radiation*, Natl Research Council Publ No 452, 1956

‡ From N Wald, Leukemia in Hiroshima city atomic bomb survivors, *Science*, 127 699–700, 1958

types of cancer have been induced experimentally, and it is clear that all types of ionizing radiation share cancer forming potency The radiation induced cancers that have been noted most often in man are cancer of the skin, leukemia, cancer of bone, and cancer of the lung

Cancer of the skin was encountered early in this century as a complication of radiation dermatitis, developing on the hands and fingers of many of the pioneer radiologists This effect resulted from prolonged manipulation of radiation equipment, the hands being exposed to large cumulative doses of radiation in the era before the attendant hazards were suspected Although this disease claimed the lives of more than 100 of the first workers in radiology, because of present safeguards it is no longer an occupational threat

Another malignant disease encountered with unusual frequency in radiologists is leukemia This disease is also abnormally prevalent in other populations exposed to radiation, such as the Japanese atomic bomb survivors

Radiation induced bone cancer was first noted in painters of luminous watch dials, who inadvertently ingested toxic quantities of radium containing paint by habitually drawing their paint brushes to a point between their lips Deposition of the radium in the skeleton, where it delivered relatively large doses of radiation to small foci of bone led to the development of skeletal cancer Osseous tumors are also abnormally prevalent in those who have consumed radium for medicinal purposes

Cancer of the lung in miners employed in the pitchblende mines of Saxony, traditionally the predominant cause of death among these workers, has been attributed to their prolonged exposure to radon which is present in high concentrations in the air of such mines (7×10^{-11} to 7×10^{-9} curie per liter) It is noteworthy, however, that many other complicating factors present in this population may have contributed to the effects of radiation in causing lung cancer

Cancer incidence versus radiation dose. Although radiation increases the incidence of many types of cancer, the precise relation between incidence and dose is not known for any type of cancer, especially at low radiation levels The yield of tumors per unit dose cannot, therefore, be accurately estimated, nor can it be assumed that any amount of radiation, however small, will increase the cancer rate. Neither can it be assumed that a threshold does exist below which no cancers are induced Since, however, for most experimentally induced cancers the incidence is not linearly proportional to dose, the evidence suggests that at very low dose levels radiation is appreciably less effective than at high dose levels, if it is effective at all Paradoxically, irradiation actually reduces the frequency of certain types of tumors in experimental animals Accordingly, no generalization about the relation between cancer incidence and radiation dose can be applied to all types of cancer

Effect of constitutional variables. Susceptibility to cancer induction in experimental animals varies widely with constitutional variables, such as genetic inheritance, age at irradiation, sex, and hormonal activity Furthermore, the influence of any one of these variables on susceptibility to tumor formation often differs from the influence of others and from one type of tumor to another As yet, little is known about the action of these variables on susceptibility in experimental animals and almost nothing about their action in man

Mechanisms of cancer induction The essential change caused by radiation that leads ultimately to the development of cancer is unknown, as is the

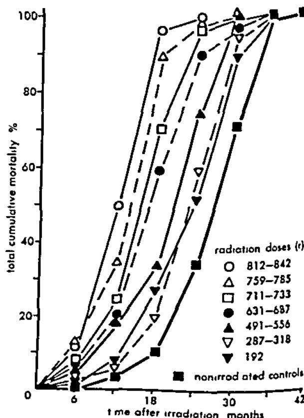


Fig 3. Longevity of female mice surviving 30 days after exposure to atomic bomb γ-rays at 6–12 weeks of age (From A C Upton, J Gerontol, 12 307–313, 1957)

death is not known but it is noteworthy that the process of photosynthesis seems *relatively* insensitive to radiation Many of *the effects of radiation* in the growth of higher plants can be attributed to destruction *of growth hormone* or depression of its synthesis

Induction of mutations by irradiation *of* seed has been used to good advantage in *breeding new* varieties of plants, *however it must be remembered that the mutations* so induced occur more or less at random But since the great majority of mutations are deleterious one improved plant is gained only at the cost of thousands of others. *The* value of the process *lies in* the ultimate derivation of a more desirable strain or species from *one* superior plant

[illegible] AFFECTING THE RADIATION RESPONSE

[illegible]

explained although a number of factors [illegible] to influence radiosensitivity Of these perhaps *the most important is position on* the phylogenetic scale for example more than 1 000 000 rad are required to kill certain bacteria whereas less than 1000 rad kill most mammals and less than 100 rad are lethal for many types of mammalian cells (Table 2)

Another determinant of radiosensitivity is the number of sets of chromosomes present in the cell This is logical since each set of chromosomes contains *genes duplicating those of another* Hence, in any given family of cells those with multiple sets of chromosomes are more resistant than those with only a single set The radiosensitivity of the cell also varies with the stage in the division cycle at which it is irradiated as well as with its rate of division, as mentioned earlier Similarly the radiosensitivity of the organism varies at different stages in its development

In addition to the aforementioned intrinsic factors affecting radiosensitivity extrinsic variables such as temperature moisture light and oxygen tension influence susceptibility Cells irradiated in the frozen state are in general relatively radioresistant presumably because of the reduced formation and diffusion of radicals at low temperature The influence of temperature is complex however it varies with the time the *temperature* is altered in relation to irradiation and with the type of cells studied In most cells vulnerability to radiation varies with oxygen tension presumably because oxidizing *radicals such as hydrogen* peroxide are important in the production of radiation injury Oxygen also exerts other effects however since in some types of cells the healing of broken chromosomes requires *energy obtained* through oxidative pathways

Physical factors of radiation The biological effects of radiation depend not merely on the amount of radiant energy *absorbed but also* on the distribution *of the radiation in* time (dose rate, or radiation intensity) and space (linear ion density or linear energy *transfer)* Certain effects such as chromosomal rearrangements require more than one ionizing particle or radical to act simultaneously within a small volume of *the cell For* multi[illegible] densely ionizing *radiations*

[illegible]

The dose rate may influence the effects of radiation in other ways too since some *types of injury undergo repair with* the passage of time For *such injuries a given* total dose is less effective if fractionated into successive widely separated increments than if administered in a single *brief* exposure Paradoxically however fractionation in some instances increases the effectiveness of radiation if the successive doses are of appropriate size and periodicity The influence of fractionation depends therefore on many variables *including* the total dose total duration of irradiation number of exposures *dose per exposure* dose rate per exposure *interval between* exposures and recovery capacity of the system irradiated

Indirect biological effects of radiation In addition to affecting the cells irradiated radiation in *directly* alters neighboring or distant cells These *indirect effects may* conceivably result from toxic substances liberated by the dying cells or from reactive changes occurring as part of the body's adaptation to injury

Thus far *attempts to demonstrate the liberation* of toxic materials from irradiated cells have yielded inconsistent results Hence *it is not established* whether such materials are produced in significant quantities It would appear however that they are rarely if ever *of major importance in the* reaction to ionizing radiation

On the other hand indirect effects resulting from adaptive or reparative processes are well documented *although* their relative importance is not always known These *include alterations caused* by local inflammation scarring and occlusion of blood vessels in addition to constitutional changes such as immunological depression hormonal disturbances and debilitation

Modification of radiation injury *It has long been* the goal of radiobiologists and radiotherapists to be able to increase or decrease radiosensitivity at will Radiotherapists have sought ways of increasing destruction of cancer cells without *damaging* normal tissues Although their objective is yet to be attained methods for modifying radiosensitivity are now known and are being tested at the experimental and clinical levels

One of these methods consists in administering drugs that reduce the effectiveness of radiation principally by blocking or inactivating radiation induced radicals The most potent of such drugs yet discovered are cysteine and *its derivatives* some of which lower the effectiveness of *x radiation in* microorganisms and laboratory animals by a factor of two or more Although most of the chemicals of this type thus far tested have proved too toxic for

however, tend to manifest themselves long after birth

GENETIC EFFECTS OF RADIATION

Since radiation induced mutations may alter characteristics inherited through eggs or sperm and may be passed on through successive generations, such effects are of greater potential significance than those affecting only the exposed individual As yet, although there is virtually no information about the genetic effects of radiation in human beings, certain facts established by investigations with other organisms appear applicable to all species, including man

The natural mutation rate. Through mechanisms yet to be discovered, genes are altered at random during the course of time These alterations vary from gross chromosomal changes, affecting many genes, to changes affecting only one gene at a time (point mutations) The former are, in general, more deleterious in their effects than the latter, but even the latter are preponderantly detrimental and, in certain instances, lethal Since genes are present in homologous pairs and one gene of each pair is inherited from each parent, the effects of a mutation depend on whether the mutant gene behaves as dominant or recessive or interacts with other genes In man, certain dominant mutations are estimated to appear naturally in 4–40 of every million eggs or sperm

Radiation induced mutations. As previously noted certain point mutations seem to result from only one ionization of the gene For a given dose of radiation however, the yield of detectable mutations in germ cells varies from gene to gene and the relative frequencies of the various mutants differ from those obtained with mutagenic chemicals suggesting that genes differ in their susceptibility to mutation depending on the nature of the mutagenic agent to which they are exposed Because of this variation and because mutations have been observed thus far in very few of the thousands of genes present in any cell only crude estimates can be made of the radiation effects on the over all mutation rate

In general the frequency of point mutations in most types of irradiated cells varies in direct proportion to the dose regardless of the dose rate Hence it has been thought that any amount of radiation is mutagenic and that the effects of successive exposures are entirely cumulative Experiments with mice however suggest that the latter may not be true of the complex populations of germ cells in mammals These experiments disclose that fractionated or chronic irradiation causes fewer detectable mutations in germ cells than brief single exposures and that the yield of mutations diminishes at very high dose levels of 1000 roentgens (r) Because of these discrepancies between mice and lower forms, any estimates of the mutation inducing effectiveness of radiation in man are of necessity highly speculative Such estimates have, nonetheless been made in an effort to assess the genetic hazards of fallout radiation, the mutation rate doubling dose for man being considered to be between 10 and 100 rad

Practical significance of genetic effects. Until more is known about the extent to which the mind and body are normally handicapped by unfavorable genes, the effects of added deleterious mutants may be only guessed at Consequently, even if the mutagenic effectiveness of radiation were known accurately, the diverse biological and social consequences of a given radiation induced increase in the mutation rate could not be estimated at present It is generally accepted, however, that the genetic burden is regulated by the rates at which mutant genes are produced and subsequently eliminated from the population through natural selection Hence it is thought that any increase in the mutation rate above natural levels augments the genetic burden

In the absence of more adequate knowledge of human genetics, attempts have been made to assess the genetic burden in terms of the morbidity from specific genetic traits, such as albinism achondroplasia, and aniridia Of such traits which are detectable in about 4% of all live births, only about one-fourth appear to be attributable to simple genetic mechanisms such as a single mutant gene Accordingly if the mutation rate were doubled by radiation although the latter group would presumably be doubled, the frequency of the other groups would not be increased so greatly The total incidence then of all such traits, although elevated from approximately 4% to a level above 5%, would probably still remain below 8%, or a value twice normal

In addition to the aforementioned traits however the occurrence of which is determined by all-or none mechanisms there are characteristics that appear to be quantitatively influenced by heredity

diation effects in man would be of the utmost gravity

EFFECTS OF RADIATION ON PLANTS

Plant cells have been used extensively for studying the genetic effects of radiation and for investigating the effects of radiation on chromosomes and on cell division Relatively little work has been done in plants however on radiation injury at the tissue or organ level From available evidence, it appears that radiation effects on plant and animal tissues are not qualitatively different if allowances are made for physiological discrepancies between the two types of organisms

Irradiation has produced a wide variety of disturbances in plants These vary from subtle changes resulting from mutations in seeds to marked inhibition of growth and to killing The cause of

diation ordinarily existing in the environment, since any additional exposure will presumably be added to that already present

Owing to concern about the potential danger of fallout from weapon tests, systematic studies have been made to ascertain the amount of radiation to which mankind is now exposed This radiation comes from several sources: (1) naturally occurring radioelements present in the earth, atmosphere, and human bodies, (2) cosmic rays, (3) man- ... x ray machines, watch dials,

materially increased by man made ... ularly medical exposure, however, the contribution from weapon fallout is almost negligible to date, despite its notoriety It is also evident from these data that the average cumulative dose received by the germ cells during the first 30 years of life is in the neighborhood of 3–8 rad

Although, as mentioned earlier, the biological effects of small doses of radiation cannot yet be estimated accurately, certain recommendations have

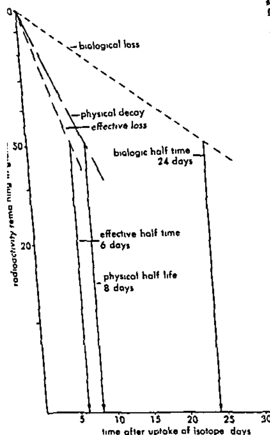


Fig 5 Rate of loss of radioactive iodine 131 after maximal uptake in the thyroid gland The rate of biological loss is plotted as exponential, although it is actually somewhat more rapid initially than at later times after uptake (From G A Andrews, Ann Internal Med, 47 931, 1957)

Table 5. Average dose of radiation to soft tissues and gonads from environmental sources*

Radiation source	Dose to gonads per year,† rad
Natural background	
External radiation	
Cosmic rays	0 028
Local γ-rays	0 047
Radon in air	0 001
Internal radiation	
Potassium-40	0 019
Carbon-14	0 001
Radon and disintegration products	0 002
Man made contributions	
Medical radiology	0 100
Shoe-fitting fluoroscopic machines	0 001
Luminous watch and clock dials	0 001
Occupational exposure	0 002
Television sets	0 001
Fallout from weapon tests	0 001
Total	0 204

... fects of ... Official ... 1958, a ... to Mar. ...

† Crude approximation to average estimate ... sea level in United States and Europe, includes allowance for relative biological effectiveness of heavy particles.

been made in an effort to safeguard man against exposure to levels of radiation likely to cause de... ...table injury Of the various possible types of

whereas somatic effects occur only in exposed subjects The Committee on Genetics of the National Academy of Sciences (1956) has therefore recommended that the maximum permissible gonadal radiation dose for the general population not exceed 10 rad during the first 30 years of life, this dose being below that estimated to double the natural mutation rate (10–100 rad) Since persons occupationally exposed to radiation (radiologists and atomic energy workers, for example) constitute but a negligible percentage of the total population, it has ... gonads

tion injuries such as life-shortening and cancer,

titative relation between these effects and dose, the risks of low level exposure cannot be estimated It is noteworthy, however, that, on the basis of extrapolation from the experimental and clinical data now available, natural background radiation cannot conceivably account for more than a small fraction of the spontaneous incidence of cancer and degenerative changes associated with senescence

See RADIOACTIVE FALLOUT [A C U]

Bibliography: G A Andrews, A few notions involved in the clinical use of radioisotopes, *Ann. In*

human use, a few of the newer agents appear promising

The opposite approach, enhancement of the effectiveness of radiation, is likewise still in the experimental stage Perhaps the most notable of the radiosensitizing agents studied is oxygen Irradiation of cancer under increased oxygen tension has been prompted by the fact that many tumors are relatively poorly oxygenated in contrast to normal tissues and to that extent relatively radioresistant

Among other modifying agents under investigation are chemicals such as nitrogen mustard and urethane which possess radiomimetic, or radiation like, activity These compounds, which are effective by themselves in curbing the growth of cancer cells, exert additive or synergistic effects when administered concomitantly with radiation

In addition to attempts at preventing or modifying radiation injury by certain processes applied before or during exposure, trial treatments after irradiation are being carried out Some cellular effects hitherto considered irreversible may be inhibited if appropriate measures are taken promptly enough after irradiation A particularly notable advance is the discovery that otherwise lethal injury of the bone marrow may be repaired by transplantation of nonirradiated blood forming cells Unfortunately, however, since immunological reactions usually occur when the donor of the marrow is genetically different from the recipient, use of this method in man will be greatly restricted until this complication can be overcome

EFFECTS OF INTERNALLY DEPOSITED RADIOELEMENTS

The radiation injuries caused by radioisotopes within the body are basically no different from those caused by radiation penetrating from without Owing to inhomogeneities of distribution, however, and other variables peculiar to internally deposited emitters, the effects of radioisotopes deserve to be considered separately

Anatomic distribution of isotope. In distribution of injury, the area damaged depends on the localization of the radioelement in the organism and on the energy of the emitted radiation

The localization of the isotope is determined by its chemical nature, physical properties and route of entry into the body Radioactive iodine 131, for example, behaves chemically like stable iodine, being concentrated from the blood stream in the thyroid gland Strontium 90 on the other hand, behaves chemically like calcium being deposited primarily in the mineral salts of bone In contrast to these two elements, which become localized in specific organs if dissolved in the blood stream, tritium like hydrogen, permeates the entire body If, conversely, these same elements are incorporated into the body as insoluble particulates their distribution is altered in that they are then concentrated in phagocytic cells, the location of which depends on the

—whatever its [illegible]

Table 4 Distribution and excretion of some radioactive fission products and other radioelements*

Radioelement	Principal organ of deposition	Half life, days	
		Physical	Biological
Barium 140	Bone	12 8	200
Calcium-45	Bone	152	18,000
Carbon 14	Bone	2 09 × 10^6	180
Cesium 137	Muscle	12 000	17
Cobalt-60	Spleen	1 900	9
Iodine-131	Thyroid	8	180
Plutonium 239	Bone	8 8 × 10^6	43 000
Polonium 210	Spleen	183 3	5
Radium 226	Bone	5 9 × 10^5	16,000
Strontium 90	Bone	9 100	3,900
Uranium 233	Bone	5 9 × 10^7	3 0
Xenon 133	Body	5 3	0 1

* Modified from Subcommittee on Permissible Internal Dose, *Maximum Permissible Amounts of Radioisotopes in the Human Body and Maximum Permissible Concentrations in Air and Water*, Handbook No 52 U S National Bureau of Standards, 1953, and J G Hamilton The metabolism of the fission products and the heaviest elements, *Radiology*, 49 328–343, 1947

ular, with the result that its radioactivity tends to be concentrated in foci For this reason and since the intensity of emitted radiation diminishes with the square of the distance from the radioactive atom, the distribution of radiation damage tends to be patchy, especially when the emission consists of α particles or low energy β particles. On the other hand, if the emission consists of penetrating γ rays, distant organs that contain no radioisotope may be injured

Elimination of isotope from the body. The degree of injury caused by a given radioelement depends on the amount of radiation it delivers to surrounding tissue This, in turn, depends on the disintegration rate of the isotope (physical decay) and on its rate of elimination from the tissue by metabolism or excretion Since the physical decay of an isotope occurs at an exponential rate and since for practical purposes its elimination from the body also occurs at a more or less exponential rate after its initial distribution, the amount remaining in the body at any given time (the "effective" concentration) will vary as the resultant of the physical and biological half lives There is of course, no consistent correlation between the two half lives, both types of values vary enormously from one isotope to another In general however, elements that are concentrated in bone tend to be eliminated very slowly and are therefore particularly hazardous Included in this group are many heavy elements and fission products

PRACTICAL HAZARDS OF IONIZING RADIATION

It is evident from the foregoing comments that, barring occasional massive exposures from nuclear accidents or atomic warfare the hazards of radiation result from small doses accumulated over a long period of time The effects of such doses fall into two categories, genetic effects and delayed somatic effects To evaluate these hazards in proper perspective, note must be taken of the level of ra

any ...
to that already present

Owing to concern about the potential danger of fallout from weapon tests, systematic studies have been made to ascertain the amount of radiation to which mankind is now exposed This radiation comes from several sources (1) naturally occurring radioelements present in the earth atmosphere, and human bodies, (2) cosmic rays, (3) man made devices such as x ray machines, watch dials, nuclear weapons, television apparatus. It is now apparent that the natural background level has been materially increased by man made radiation, particularly medical exposure, however, the contribution from weapon fallout is almost negligible to date ... notoriety It is also evident from these

Although as mentioned earlier, the biological effects of small doses of radiation cannot yet be estimated accurately certain recommendations have

Table 5 Average dose of radiation to soft tissues and gonads from environmental sources*

Radiation source	Dose to gonads per year, rad
Natural background	
External radiation	
Cosmic rays	0 028
Local γ rays	0 047
Radon in air	0 001
Internal radiation	
Potassium-40	0 019
Carbon 14	0 001
Radon and disintegration products	0 002
Man made contributions	
Medical radiology	0 100
Shoe fitting fluoroscopic machines	0 001
Luminous watch and clock dials	0 001
Occupational exposure	0 002
Television sets	0 001
Fallout from weapon tests	0 001
Total	0 204

sea level in ...
for relative biological effectiveness of heavy particles

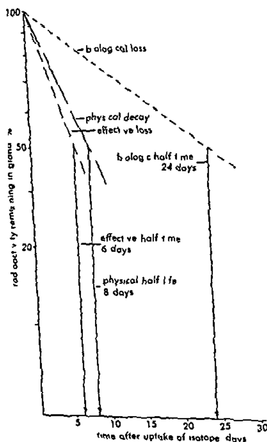


Fig 5 Rate of loss of radioactive iodine 131 after maximal uptake in the thyroid gland The rate of biological loss is plotted as exponential although it is actually somewhat more rapid initially than at later times after uptake (From G A Andrews Ann Internal Med., 47 931, 1957)

been made in an effort to safeguard man against exposure to levels of radiation likely to cause detectable injury Of the various possible types of injury, genetic effects have been considered more important than somatic effects since they are essentially irreversible and threaten future generations, whereas somatic effects occur only in exposed subjects The Committee on Genetics of the National Academy of Sciences (1956) has therefore recommended that the maximum permissible gonadal radiation dose for the general population not exceed 10 rad during the first 30 years of life, this dose being below that estimated to double the natural mutation rate (10-100 rad) Since persons occupationally exposed to radiation (radiologists and atomic energy workers for example) constitute but a negligible percentage of the total population it has been recommended that the dose to their gonads may be somewhat higher but that it still not exceed 50 rad before age 30

Concerning the hazard of delayed somatic radiation injuries such as life shortening and cancer, there is no conclusive evidence as yet that doses only slightly above the natural background are damaging Hence until more is known about the quantitative relation between these effects and dose the risks of low level exposure cannot be estimated It is noteworthy, however, that, on the basis of extrapolation from the experimental and clinical data now available, natural background radiation cannot conceivably account for more than a small fraction of the spontaneous incidence of cancer and degenerative changes associated with senescence See RADIOACTIVE FALLOUT [A C U]

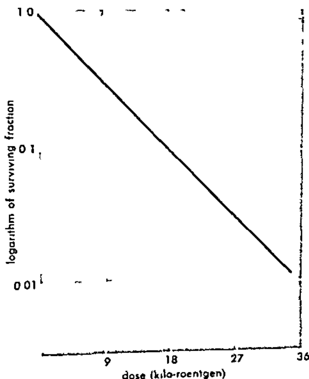
Bibliography G A Andrews A few notions involved in the clinical use of radioisotopes, *Ann In*

ternal Med, 47 922–938, 1957, C F Behrens (ed), *Atomic Medicine*, 2d ed, 1953; A E Hollaender (ed), *Radiation Biology*, vol 1, 1954, D E Lea, *Actions of Radiations on Living Cells*, 1947, E B Lewis, Leukemia and ionizing radiation, *Science*, 125 965–972, 1957, National Academy of Sciences, *The Biological Effects of Atomic Radiation*, Natl Research Council Publ, 1956, A H Sparrow, S A Gordon, K Sax, C F Konzak and J E Gunkel Symposium on the effects of ionizing radiation on plants, *Quart Rev Biol*, 32(1).1–56, 1957, A C Upton, Ionizing radiation and the aging process, *J Gerontol*, 12 306–313, 1957, A C Upton, Studies on the mechanism of leukemogenesis by ionizing radiation in *Carcinogenesis Mechanisms of Action*, Ciba Foundation Symposium, in press, A C Upton F P Conte G S Hurst, and W A Mills, The relative biological effectiveness of fast neutrons X rays, and α rays for acute lethality in mice, *Radiation Research*, 4 117–131, 1956

Radiation microbiology

A field of basic and applied radiobiology concerned chiefly with the damaging effects of radiations on microorganisms The basic studies are usually designed to elucidate the mechanism by which radiations produce their biological damage The applied field, radiation food sterilization, has as its ultimate goal the "cold" sterilization of food and food products by subjecting them to large doses of ionizing radiation This article will deal with effects of ionizing radiation See ULTRAVIOLET RADIATION BIOLOGY

Basic radiation microbiology. Viruses algae, bacteria yeast and other fungi as well as protozoa have been used as test objects for radiation studies



X ray dose effect curve on bacteria

This group represents a broad range of size and biological complexity The size and homogeneity of the populations available, and the simplicity of technique in growing, harvesting and quantitative measurement of radiation damage in some of these organisms, especially free living cells, have been deciding factors in the choice of test object A number of radiation effects have been studied on these cells, including lethal effects, mutation and alterations in the biochemical and physiological activity See ALGAE, BACTERIA, FUNGI, VIRUS YEAST

Lethal effects These are often measured by the loss of reproductive capacity of the irradiated cells This effect is quantitated in most experiments by determining the reduction in number of irradiated cells capable of forming macroscopically visible colonies on sterile nutrient solid medium Formation of such colonies results from a large number of cell divisions and does not indicate when the cells die Killing by ionizing and ultraviolet radiations is a simple function of dose for a number of microorganisms

Mutagenic effects Radiation induced mutations are often measured in microorganisms by determining the absolute number of cells in an irradiated population that have lost the ability to perform a specific step in a biochemical synthetic pathway or have lost or gained resistance to an antibiotic substance or a virus These alterations to be classified as changes in the hereditary material should be maintained in subsequent generations of the living cell in question A difficulty in determining the frequency of such events in microorganisms is that in an irradiated population of cells such changes are accompanied by lethality

Physiological and biochemical changes A number of radiation induced changes occur in microorganisms for example alterations in the permeability of the cell membranes loss of synthetic capacity delay in cell division and induced hypersensitivity to other physical and chemical agents Some of these effects are now known to be reversible

Target theory considerations The similarity between the randomness of the physical events known to occur along the track of an ionizing particle and the randomness of occurrence of effects in a microbial population was recognized by early radiobiologists The hit theory for production of biological damage resulted from the recognition of this interrelation Consideration of the nature of the physical events that is an ionization occurring in or passage of an ionizing particle through a vital and structure of the target could be deduced from proper radiation studies

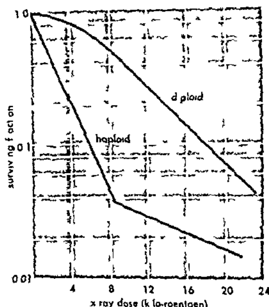
Nature of the target The assumption inherent in the target hypothesis is that the occurrence of the physical event within a subcellular structure leads to the effect In other words, a structure

considered to be the target Analysis of survival curves for some microorganisms indicates that the sensitive volume is less than 0 001 of the cell vol ume If a spherical shape is assumed the target can be deduced to have a diameter of approxi mately 200 angstroms There are many structures of he e d mensions within the cell Most of the

produce the effects It seems reasonable therefore to assume that the sensitive material must be pres ent within the cell in limiting amounts and that the material must be crucial to the continuity of

However the same argument can be made for any vital enzyme or enzyme system present within the cell in limiting concentration Several lines of evidence seem to implicate the genetic material in the lethal effect In microorganisms that have a well defined nucleus as well as in cells of higher forms of life irradiation of the nucleus is more effective in killing than irradiation of the cyto plasm or nonnuclear portion There are strains of yeast available for investigation that have discrete multiples of chromosomal sets and therefore mul tiple sets of identical genes (see POLYPLOIDY) Haplo d cells which have one set diploid cells which have two sets and higher ploidy series of cells have been studied to determine the relative radiosensitivity as a function of the multiplicity of genetic material The results of such studies show that diploid cells are more resistant than haplo d cells which suggests that radioresistance is a function of the degree of ploidy Unfortunately the higher ploidy series do not show a simple rela tion between radioresistance and ploidy In some cells of higher forms of life the size of the cell parallels the degree of ploidy On this basis it might be assumed that the enzyme content in creases with increasing ploidy These results are suggestive as far as mechanism is concerned but do not prove that killing is caused by lethal muta tions In bacteria evidence for genetic killing is far less convincing than in the microorganisms al ready mentioned See MUTATION RADIATION IN JURY (BIOLOGY)

Events leading to biological effect Much interest has been aroused by the large number of studies on mod fication of the effects of ionizing radiation by environmental factors Results of these studies ind cate the possibility of elucidating the nature of chemical events occurring between the initial or primary physical processes and biological effect. The finding that several relatively simple environ mental factors can profoundly change the efficiency of the radiation in producing its effect suggest that i n zing radiation may act by production of toxic ox dizing free radicals or molecules in cells in a normal physiological environment or by activation of the target molecule which is then inactivated



Survival of haplo d and d plo d yeast cells (R Latar jef and B Ephrussi in Compt rend 229 306 1949)

by further reaction with oxygen or another acti vated molecule or by a normal molecule in the vicinity of the target The table shows some of the factors that can alter the efficiency of x or γ rays in inactivation of a bacterium

Some of these conditions such as oxygen removal and freezing are known to alter the yield of oxi dizing radiodecomposition products in irradiated water Although the effects in the biological system and in pure water are similar the results do not necessarily prove that free radicals and hydrogen peroxide are the toxic agents responsible for cell inactivation

Postirradiation conditions can also modify the effect produced per unit dose The temperature at which irradiated bacteria are incubated and the presence or absence of certain nutritional factors in the medium in which the cells are incubated influence the effectiveness of the radiation. The response of irradiated cells to these postirradiation treatments has been demonstrated both for ionizing radiations and for ultraviolet radiation Results of experiments that involve modifying treatments do

Mod fication of the bactericidal effect of x rays on *Escherichia coli* B/r by environmental factors

Treatment before and during irradiation	Protective factor*
Oxygen removal	3
Chemical treatment	
Cysteine	2 5
2,3-D mercaptopropanol (0 02 M)	4
Sodium hydrosulfite (0 02 M)	4
β-Mercaptoethylamine (0 04 M)	8 12
Freez ng (oxygen present)	5–6
Freezing (oxygen absent)	10–12

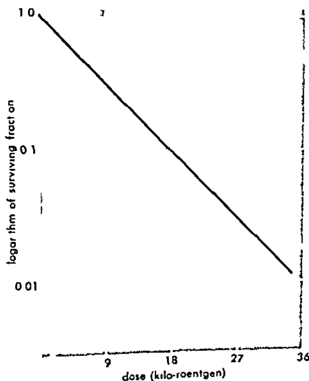
* The ratio of dose requ red to produce the same per centage killing with the procedure used as compared to a standard suspens on irradiated in the presence of air

ternal Med 47 922–938 1957 C F Behrens (ed) *Atomic Medicine* 2d ed 1953, A E Hollaender (ed) *Radiation Biology* vol 1 1954 D E Lea *Actions of Radiations on Living Cells* 1947 E B Lewis Leukemia and ionizing radiation *Science,* 125 965 972 1957 National Academy of Sciences *The Biological Effects of Atomic Radiation* Natl Research Council Publ 1956 A H Sparrow S A Gordon K Sax C F Konzak and J E Gunkel Symposium on the effects of ionizing radiation on plants *Quart Rev Biol* 32(1) 1–56 1957, A C Upton Ionizing radiation and the aging process *J Gerontol* 12 306–313 1957 A C Upton Studies on the mechanism of leukemogenesis by ionizing radiation in *Carcinogenesis Mechanisms of Action* Ciba Foundation Symposium in press, A C Upton F P Conte G S Hurst and W A Mills The relative biological effectiveness of fast neutrons X rays and α rays for acute lethality in mice *Radiation Research* 4 117–131 1956

Radiation microbiology

A field of basic and applied radiobiology concerned chiefly with the damaging effects of radiations on microorganisms The basic studies are usually designed to elucidate the mechanism by which radiations produce their biological damage The applied field radiation food sterilization has as its ultimate goal the cold sterilization of food and food products by subjecting them to large doses of ionizing radiation This article will deal with effects of ionizing radiation *See* ULTRAVIOLET RADIATION BIOLOGY

Basic radiation microbiology Viruses algae bacteria yeast and other fungi as well as protozoa have been used as test objects for radiation studies



X ray dose effect curve on bacteria

This group represents a broad range of size and biological complexity The size and homogeneity of the populations available, and the simplicity of technique in growing harvesting and quantitative measurement of radiation damage in some of these organisms especially free-living cells, have been deciding factors in the choice of test object A number of radiation effects have been studied on these cells including lethal effects, mutation, and alterations in the biochemical and physiological activity *See* ALGAE BACTERIA FUNGI VIRUS YEAST

Lethal effects These are often measured by the loss of reproductive capacity of the irradiated cells This effect is quantitated in most experiments by determining the reduction in number of irradiated cells capable of forming macroscopically visible colonies on sterile nutrient solid medium Formation of such colonies results from a large number of cell divisions and does not indicate when the cells die Killing by ionizing and ultraviolet radiations is a simple function of dose for a number of microorganisms

Mutagenic effects Radiation induced mutations are often measured in microorganisms by determining the absolute number of cells in an irradiated population that have lost the ability to perform a specific step in a biochemical synthetic pathway or have lost or gained resistance to an antibiotic substance or a virus These alterations to be classified as changes in the hereditary material should be maintained in subsequent generations of the living cell in question A difficulty in determining the frequency of such events in microorganisms is that in an irradiated population of cells such changes are accompanied by lethality

Physiological and biochemical changes A number of radiation induced changes occur in microorganisms for example alterations in the permeability of the cell membranes, loss of synthetic capacity delay in cell division and induced hypersensitivity to other physical and chemical agents Some of these effects are now known to be reversible

Target theory considerations The similarity between the randomness of the physical events known to occur along the track of an ionizing particle and the randomness of occurrence of effects in a microbial population was recognized by early radiobiologists The hit theory for production of biological damage resulted from the recognition of this interrelation Consideration of the nature of the physical events that is an ionization occurring in or passage of an ionizing particle through a vital structure resulted in the effect under investigation led to the formulation of the target hypothesis The hypothesis led to the conclusion that the dimensions and structure of the target could be deduced from proper radiation studies

Nature of the target The assumption inherent in the target hypothesis is that the occurrence of the physical event within a subcellular structure leads to the effect In other words a structure is

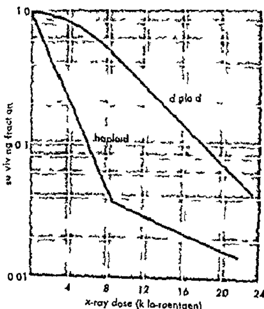
considered to be the target Analysis of survival curves for some microorganisms indicates that the sensitive volume is less than 0 001 of the cell volume If a spherical shape is assumed the target can be deduced to have a diameter of approximately 200 angstroms There are many structures of these dimensions within the cell Most of the

altered by

produce the effects. It seems reasonable therefore to assume that the sensitive material must be present within the cell in limiting amounts and that

cell in limiting concentration Several lines of evidence seem to implicate the genetic material in the lethal effect In microorganisms that have a well defined nucleus, as well as in cells of higher forms of life, irradiation of the nucleus is more effective in killing than irradiation of the cytoplasm or nonnuclear portion There are strains of yeast available for investigation that have discrete multiples of chromosomal sets and therefore multiple sets of identical genes (see POLYPLOIDY) Haploid cells which have one set diploid cells, which have two sets and higher ploidy series of cells have been studied to determine the relative radiosensitivity as a function of the multiplicity of genetic material The results of such studies show that diploid cells are more resistant than haploid cells which suggests that radioresistance is a function of the degree of ploidy Unfortunately the higher ploidy series do not show a simple relation between radioresistance and ploidy In some cells of higher forms of life the size of the cell parallels the degree of ploidy On this basis it might be assumed that the enzyme content increases with increasing ploidy These results are suggestive as far as mechanism is concerned but do not prove that killing is caused by lethal mutations In bacteria evidence for genetic killing is far less convincing than in the microorganisms already mentioned See MUTATION RADIATION INJURY (BIOLOGY)

Events leading to biological effect Much interest has been aroused by the large number of studies on modification of the effects of ionizing radiation by environmental factors Results of these studies indicate the possibility of elucidating the nature of chemical events occurring between the initial or primary physical processes, and biological effect The finding that several relatively simple environmental factors can profoundly change the efficiency of the radiation in producing its effect suggest that ionizing radiation may act by production of toxic oxidizing free radicals or molecules in cells in a normal physiological environment or by activation of the target molecule, which is then inactivated



Survival of haploid and diploid yeast cells (R Latarjet and B Ephrussi in Compt rend., 229 306 1949)

in inactivation of a bacterium

Some of these conditions such as oxygen removal and freezing are known to alter the yield of oxidizing radiodecomposition products in irradiated water Although the effects in the biological system and in pure water are similar the results do not necessarily prove that free radicals and hydrogen peroxide are the toxic agents responsible for cell inactivation

Postirradiation conditions can also modify the effect produced per unit dose The temperature at which irradiated bacteria are incubated and the presence or absence of certain nutritional factors in the medium in which the cells are incubated influence the effectiveness of the radiation The response of irradiated cells to these postirradiation treatments has been demonstrated both for ionizing radiations and for ultraviolet radiation Results of experiments that involve modifying treatments do

Modification of the bactericidal effect of x rays on Escherichia coli B/r by environmental factors

Treatment before and during irradiation	Protective factor*
Oxygen removal	3
Chemical treatment	
Cysteine	
2,3 D mercaptopropanol (0 0? M)	2 5
Sodium hydrosulfite (0 02 M)	4
β Mercaptoethylamine (0 04 M)	4
Freezing (oxygen present)	8 12
Freezing (oxygen absent)	5–6
	10–12

* The ratio of dose required to produce the same percentage killing with the procedure used as compared to a standard suspension irradiated in the presence of air

not pinpoint the nature of the reactions, but they do indicate that there are a number of reaction steps between the physical event and the irreversible fixing of the biological damage The success of postirradiation treatments in reducing the measurable damage also indicates that reasonably long times are required for establishment of the damage

Applied radiation microbiology. Food technologists now take advantage of high intensity radiation sources to remove microbiological contamination of foodstuffs Heat labile drugs are routinely sterilized by high energy electrons A variety of foods and food products can be essentially freed of bacteria by bombardment with large doses of γ rays in the absence of heat (cold sterilization) One of the disadvantages of this process is the production of off flavors in food products that contain a high concentration of fatty acids Some of the conditions that will prevent the off flavors also reduce the lethal effectiveness of the radiation Spores of some bacteria are relatively resistant to ionizing radiation, and their elimination from the food determines the length of exposure required Progress in this field is dependent to a great extent on information from basic studies in radiation microbiology See FOOD PRESERVATION, RADIATION BIOCHEMISTRY, RADIATION BIOLOGY [G E S]

Bibliography M Bacq and P Alexander, *Fundamentals of Radiobiology*, 1955, R S Hannon, *Research on Science and Technology of Food Preservation by Ionizing Radiations*, 1956, A Hollaender (ed), *Radiation Biology*, vol 2, 1955, D E Lea *Actions of Radiations on Living Cells*, 2d ed , 1955

Radiation pressure

Electromagnetic radiation transmits energy, possesses momentum, and exerts a pressure on objects on which it impinges In the case of a plane electromagnetic wave incident normally on a plane absorbing sheet the mean pressure is

$$\bar{P} = \tfrac{1}{2}\epsilon E_0^2$$

where E_0 is the amplitude of the electric field, and ϵ is the dielectric constant of the medium If the wave impinges normally on a perfectly reflecting plane conducting sheet, then standing waves are formed, and the average pressure is twice that on the absorbing sheet These pressures are very small (about 10^{-9} newton/m² if E is a few volts per meter), but were measured successfully by E F Nichols and G F Hull in 1903 The effect is conspicuous in the case of a comet near the sun where the radiation pressure from the sun forces the lighter cometary constituents away from the sun See ELECTROMAGNETIC RADIATION MAXWELL'S EQUATIONS, WAVE EQUATION [W R SM]

Radiation shielding

Material interposed between a source of radiation and a radiation sensitive body to protect the latter, also, the process of effecting such shielding Sources of radiation include radioactive isotopes, x ray machines, nuclear reactors and cosmic rays Radiation sensitive bodies to be shielded are usually people, but may include radiation detection instruments, photographic film, or materials whose physical properties are changed by radiation, such as rubber, cloth, certain chemicals, or electronic components

For the effects of radiation on living organisms see RADIATION BIOLOGY, RADIATION INJURY (BIOLOGY) For the effects on inanimate materials, see RADIATION DAMAGE (INANIMATE MATERIALS) See also NUCLEAR EXPLOSION

Damaging radiation includes charged particles (α and β rays), uncharged particles (neutrons) and electromagnetic radiation (γ rays, x rays) Because of their electric charge, α and β rays are easily stopped Neutrons, γ rays, and x rays, however, are more penetrating and present a serious shielding problem

Radiation sources The most common strong source of neutrons is the nuclear reactor in which neutrons are produced in the fission process They are distributed in energy approximately as shown in Fig 1, which shows the distribution for uranium-235 fission That for other fissionable isotopes is similar While the most probable energy is about ¾ Mev, and the mean energy about 2 Mev, the most important neutrons from reactor shielding considerations are those produced at considerably higher energy, usually about 8 Mev, because of their greater penetrating ability Additional neutrons are produced by fission product decay and charged particle reactions See FISSION, NUCLEAR, REACTOR NUCLEAR, REACTOR PHYSICS

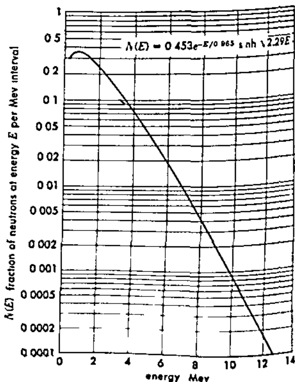
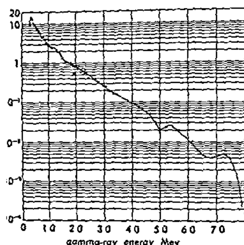


Fig 1 Neutron spectrum from the fission of U²³⁵ by thermal neutrons. (From L Cranberg et al, Phys Rev 103 662–670 1956)



2 Energy spectrum of gamma rays observed 1 n 10^{-7} sec after fiss on

The γ-rays produced in a reactor come from sev l sources γ rays emanating immediately from : fissioning nuclei are distributed in energy as own n Fig 2

Gamma rays are also produced in radioactive cay of fiss on products and of neutron activated iterials These determine shielding requirements : a reactor after shutdown

Neutron capture constitutes the single most im rtant source of γ rays in reactors The total en gy of capture γ rays from a single neutron cap re event is equal to the binding energy of the utron in the nucleus which is about 7 Mev for ost elements Notable exceptions include hydro n (2 24 Mev) and nitrogen (10 8 Mev) See AMMA RAYS NEUTRON NEUTRON CROSS SECTION

Neutrons can also cause γ ray production by in astic scattering. The neutrons must have sufficient netic energy to enable excitation to at least the west energy level above the ground state of the ruck nucleus This process while disadvantageous cause of the production of secondary γ rays is ten desired because in the process the neutron is owed down and deflected rendering it less likely penetrate the shield

Attenuation processes Gamma rays are atten ated primarily by three processes the photoelec ic effect, the pair production process and Comp n interactions The last is a scattering process, nd the first two are tantamount to absorption For etailed information of these processes see COMP ON EFFECT PAIR PRODUCTION (ELECTRON POSI RON) PHOTOEMISSION

The Compton effect is accounted for in shield ng calculations by multiplying the intensity due o uncollided photons by a buildup factor" which s very roughly equal to one plus the number of mean free paths traversed The cross sections for photoelectric effect pair production and Compton ffect vary from element to element roughly as Z^5 Z^2 and Z respectively Z being the atomic number of the shielding material The total cross section is the sum of the cross sections for the in dividual processes shown as a function of energy for lead in Fig 3 In Fig 4 are shown the total macroscopic cross sections that is the attenuation coefficients μ divided by density ρ for all ele ments and several photon energies

Neutrons are attenuated in shields primarily by being first slowed down and then absorbed Slow u ... achieved by elastic scattering from l ... inelastic ... le hydro ... all ener gies above thermal (above ... ev) the heavy elements do not scatter inelastically for neu trons below the energy threshold which varies from many Mev in some elements to about 100 kev in the best inelastic scatterers Neutron capture after the neutron is slowed down occurs within a short distance in most materials Neutron shield mate rials are chosen primarily for their ability to slow down the fast neutrons To allow for the extra shield penetration following scattering a buildup factor for neutrons is also used which is roughly equal to the exponential e with positive exponent

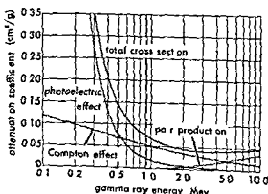


Fig 3 Gamma ray attenuat on coeff c ents for lead (After G W Grodstein)

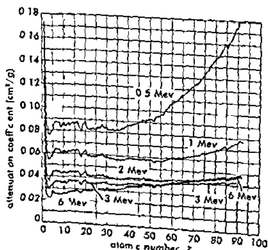


Fig 4 Gamma-ray attenuat on coeffic ents versus atom c number of absorb ng mater al for var ous en erg es

equal to one third the number of mean free paths traversed

Thus, if a point isotropic source of neutrons or γ rays is shielded by a surrounding spherical shield, the penetrating flux is given roughly by

$$\phi(R) = \frac{SB(t/\lambda)}{4\pi R^2} e^{-t/\lambda} \quad \text{cm}^{-2}\,\text{sec}^{-1}$$

where S – source strength, neutrons or photons sec^{-1}
R = distance from source to measuring point, cm
t = shield thickness, cm, with $t \leq R$
λ = mean free path of source neutrons or photons in the shield, cm, $\lambda = \mu^{-1}$ for γ rays
$B(t/\lambda)$ = buildup factor
$\cong 1 + t/\lambda$ for γ rays
$\cong e^{t/3\lambda}$ for fast neutrons

Shielding materials. The most commonly used criteria for selecting shielding materials are radiation attenuation, ease of heat removal, resistance to radiation damage, expense, and structural strength

For neutron attenuation, the lightest shields are usually hydrogenous (ammonia, lithium hydride, water), and the thinnest shields contain a high proportion of iron, copper, or other dense material For γ ray attenuation, the high atomic number elements are generally the best The production of γ rays in the shield caused by neutron capture can be suppressed by including in the shield nuclei such as those listed in the table, these absorb neutrons and release charged particles For an explanation of the reaction nomenclature used in the table, *see* NUCLEAR REACTION

For heat removal, particularly from the inner layers of a shield the material must have good heat conductivity or be capable of being circulated to a heat exchanger, that is, a coolant such as water, an organic fluid, or a liquid or molten metal (mercury, lead, bismuth, or lead bismuth alloy) Since the water organic fluid, and corrosion impurities picked up by all the coolants will become radioactive because of neutron capture shielding must also be provided around piping and the heat exchanger

Metals are the most resistant to radiation damage although there is some change in their mechan-

Nuclei useful in suppression of capture γ rays

Target nucleus	Reaction	Thermal neutron capture cross section barns*		Energy release, Mev
		Natural element	Isotope	
Li^6	$Li^6(n,\alpha)H^3$	71	945	4 78
B^{10}	$B^{10}(n,\alpha)Li^7$	44	222	2 792
	$B^{10}(n,\alpha)Li^7$ + 0 48 Mev γ	711	3591	2 792
N^{14}	$N^{14}(n,p)C^{14}$	1 75	1 75	0 624
	$N^{14}(n,\gamma)N^{15}$	0 08	0 08	10 8

* Barn = 10^{-24} cm² = measure of probability of capture

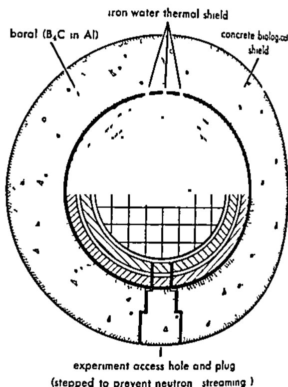


Fig 5 Typical reactor shield configuration

ical properties Concretes, frequently used because of their relatively low cost, hold up well, however if heated they lose water of crystallization, becoming somewhat weaker and less effective in neutron attenuation Although hydrogenous materials are particularly sensitive simple molecules such as water and ammonia are exceptions because of recombination of decomposition products and ease of removal of these products

If shielding cost is important, cost of materials must be balanced against the effect of shield size on other parts of the reactor facility, for example building size and support structure If conditions warrant concrete can be loaded with locally available material such as natural minerals (magnetite or barytes) scrap steel, water or even earth

Typical shields. Reactor shields vary with application and with reactor type The over all thickness of material is chosen to reduce the biological dose rate outside to a tolerable level taken to be 2.5 milliroentgen equivalent man per hour (mrem/hr) at U S Atomic Energy Commission-approved installations The level is usually made about one tenth of this or less to ensure adequate protection

regions the biological shield and the thermal shield The thermal shield located next to the reactor is designed to absorb most of the energy of the escaping radiation It is often made of steel and water and is cooled by a circulating fluid The biological shield is added outside to reduce the external dose rate to a tolerable level Figure 5 shows a typical configuration [F F B]

Bibliography H Etherington (ed), *Nuclear Engineering Handbook* 1958 H Goldstein, *Funda*

mental Aspects of Reactor Shielding, 1959, B T Price, C Horton, and K Spinney, *Radiation Shielding* 1957, T Rockwell (ed), *Reactor Shielding Design Manual*, USAEC TID 7004, 1956

adiator

f ... that
ny
nit
pace.
adiant heating, radiators are exposed to view They often heat also by conduction to the adjacent

ired by wood coal, charcoal oil or gas (such as

heater suitably circulated through the heat emitting units.

Cast iron radiators are made in sections of varying widths and heights and are assembled in the required number by top and bottom nipples Some are made in a flat panel They are set on legs or similar supports or affixed to side walls by adjustable hangers The preferred location is under windows

When windows are low cast iron baseboard radiators, less than 1 ft high are available for use

tions, extension blocks and inverted and projecting corner covers cast iron convectors may be used also

Finned tube radiators consist of a pipe or tube with affixed fins of either steel copper or aluminum They are available with or without enclosures and usually are wall mounted They are more compact than cast iron radiators having greater heating area per volume

Related devices Convectors are customarily made of nonferrous finned tubes in a wide variety of enclosures for free standing recessed or wall hung installation Air circulates over the elements by natural convection or by fans in which case they are known as unit ventilators Convectors and unit ventilators are not true radiators because they emit most of the heat by conduction to the circulated air

A limited amount of cooling may be produced by passing chilled water through radiators but care must be exercised to dispose of moisture condensation in humid weather

Electric heating elements may be substituted for the fluid heating elements in all types of radiators, convectors and unit ventilators *See* COMFORT CONTROL HOT WATER HEATING SYSTEM RADIANT HEATING STEAM HEATING [E L W]

Bibliography American Society of Heating and Air Conditioning Engineers *Heating Ventilating and Air Conditioning Guide* 1959

Radio

Communication between two or more points, employing electromagnetic waves as the transmission medium For many purposes electromagnetic waves are an ideal means of transmission, because they travel fast (186,000 miles per second), can be produced easily and economically, require only space (which is unlimited) for a medium, and can be intercepted in receivers which are small, efficient, and inexpensive

General principles. When an electric current flows through a wire, energy is stored in the form of an electric field which surrounds it If this current alternates rapidly in direction, energy in the field is converted to electromagnetic waves which can be detected at a distance They are radiated through space, as are light waves from a luminous object In practice, the wire around which the radio waves form is specially designed and elevated above interfering obstacles to function efficiently as an antenna *See* ELECTROMAGNETIC RADIATION

Radio waves and light waves are identical in composition and are identified individually only because the frequencies are vastly different and produce different effects Waves having an alternation frequency of about 15 000 or more cycles per second (cps) become useful for communications

The first significant radio applications employed frequencies of about 500 kilocycles (kc) for ship communications Their value was demonstrated dramatically in 1909 and 1912 when help was obtained by the sinking passenger ships *Republic* and *Titanic*, and hundreds of lives were saved Roughly 10 years later the modern overseas radio communications complex was born, upon discovery that frequencies of about 3000–30,000 kc traveled over great distances by reflection from the ionosphere Later many other services such as television, FM broadcasting, radar and microwave relaying, were developed as the unique properties of much higher frequencies were discovered and exploited

Information transmission. Radio waves transmitted continuously, with each cycle an exact duplicate of all others indicate only that a *carrier* is present. The message must cause changes in the carrier which can be detected at a distant receiver The carrier must be modulated by the message, such as sending it at intervals to form dots and

shown in the illustration *See* MODULATION

Methods of radio communication. Communication by radio can be effected by many different methods The method used is determined by the information to be transmitted and the purpose of the communication system

Code telegraphy The carrier is keyed on and off to form dots and dashes This technique is used in ship-to-shore communication but has been largely superseded by more efficient techniques in other services. *See* AMPLITUDE-MODULATION RADIO

Frequency shift transmission The carrier frequency is shifted a fixed amount to correspond with telegraphic dots and dashes, or with combinations of pulse signals identified with the characters on a typewriter This technique is widely used in handling the large volume of public message traffic on long circuits, principally by the use of teletypewriters *See* FREQUENCY MODULATION RADIO

Amplitude modulation The amplitude of the carrier is made to fluctuate corresponding to the fluctuations of a sound wave, a television picture signal, etc This technique is used in AM broad casting, television picture transmission, and many other services *See* AMPLITUDE MODULATION

Frequency modulation The frequency of the carrier is made to fluctuate around an average axis to correspond to the modulating wave This tech nique is used in FM broadcasting, television sound transmission, and microwave relaying A related technique, called phase modulation, is used in pub lic safety, land transportation, and many other services above 30 Mc *See* FREQUENCY MODULATION, PHASE MODULATION

Pulse transmission The carrier is transmitted in short pulses which change in repetition rate, width or amplitude or in complex groups of pulses which vary from group to succeeding group in accordance with the message information These forms of pulse transmission are identified as pulse code, pulse time, pulse position, pulse-amplitude, pulse-width or pulse frequency modulation Such techniques are complex and are employed princi pally in microwave relay systems *See* PULSE MODULATION

Radar The carrier is normally transmitted as short pulses in a narrow beam, similar to a search light When a wave pulse strikes an object such as an aircraft, energy is reflected back to the station, which measures the round trip time and converts it to distance An airborne radar can receive varying reflections from the earth beneath it and display these reflections in a maplike presentation on a cathode-ray tube *See* AIRBORNE RADAR, RADAR

Uses of radio. The first practical application of radio over a half century ago between ships and shore stations, was followed quickly by communi cation overseas and between other widely separated fixed points Subsequently the applications have become widely diversified Some of the important uses are listed below

1 Public safety Marine and aviation communi cations police and fire protection, forestry conser vation highway traffic control

2 Industrial Power utilities, pipelines, relay services, news systems agriculture, petroleum processing

3 Land transportation Railroads, motor carriers taxicabs, automobile emergency needs

4 Broadcasting Television FM broadcasting, standard broadcasting short wave international broadcasting

5 Military Radar, communications, navigation, telemetering, missile tracking, guidance, detection, and aiming

6 Fixed point to point. Long distance message and picture transmission

7. Relaying Television, sound, picture, and public message relaying over long distances.

8 Telemetering Remote indication of water levels in reservoirs and rivers, performance of experimental aircraft, missile and satellite performance, and other data

Radio channels. Hundreds of thousands of radio transmitters exist, each requiring a carrier of some frequency If all operated on the same frequency, the interference would be intolerable. This is prevented by using different carrier frequencies for transmitters where service areas overlap, and building receivers which select only the carrier frequency of the desired station Resonant electric circuits in the receiver are adjusted, or tuned, to accept one frequency and reject others

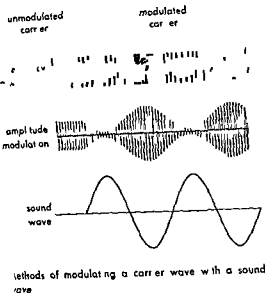
Each station operates within a specific rf channel All other stations within a geographical area are excluded from using this channel Each channel must be wide enough to accommodate the message information, provide tolerance for small carrier frequency drift, perhaps provide a guard band and allow for imperfect receiver selectivity capabilities The minimum usable channel widths (or band widths) vary from service to service, depending upon the amount of information a channel must accommodate In television it is 6000 kc, because of the large amount of essential picture information. In FM broadcasting it is 200 kc, and in AM broadcasting 10 kc The great demand for authorizations requires efficient channel utilization. In the mobile transportation service, the FCC was compelled to reduce channel widths 50% when technical developments made it feasible *See* BANDWIDTH REQUIREMENTS (COMMUNICATIONS)

FCC regulations. All nations have a sovereign right to use freely any or all parts of the radio spectrum But a growing list of international agreements and treaties divides the spectrum and specifies sharing among nations for their mutual benefit and protection

Every nation designates its own regulatory agency Functioning within the international agreements it issues authorizations, assigns frequencies, polices operations, creates technical standards, rules and practices, and safeguards and protects the public interest

In the United States all nongovernmental radio communications are regulated by the Federal Communications Commission according to the provision of the Communications Act of 1934, as amended Creation of a radio station or service requires authorization by the FCC. Upon completion of an authorized facility a license to operate is issued Radio stations are inspected regularly by engineers attached to FCC field offices Stations must comply with the terms of their authorization regarding carrier frequency tolerance, power limitations, permissible communications, call signals, and control by properly licensed personnel

Volume requirements The demand for FCC radio authorizations is incr eat fo



Methods of modulating a carrier wave with a sound wave

sting and new services In the United States alone there are nearly 2,000 000 listed authorizations for sixty different kinds of services There are over 1,500 000 outstanding radio operator licenses or permits to persons who operate authorized stations The approximate number of authorizations which have been issued for stations in various categories is shown below

AM broadcasting	3500	Common carrier	2000
TV broadcasting	900	Land transportation	310 000
FM broadcasting	800	Industrial	325 000
Public safety	250 000	Aviation	60 000
Amateur	160 000	Marine	70 000

Radio spectrum allocations The rf spectrum covers the range from about 15 000 to 100 000 000 000 cycles per second Within it are loosely divided segments which have unique propagation characteristics Frequencies from these segments are allocated to services which take advantage of their natural characteristics See RADIO SPECTRUM ALLOCATIONS RADIO-WAVE PROPAGATION [R F C U]

Bibliography W L Everitt (ed) *Fundamentals of Radio and Electronics* 2d ed 1958 K Henney *Radio Engineering Handbook* 4th ed 1958 A Hund *Frequency Modulation* 1942 Federal Communications Commission *Rules and Regulations* F E Terman, *Radio Engineers Handbook* 1943

Radio astronomy

That branch of astronomy which studies heavenly bodies by observations of the radio waves that they emit The fact that radio waves are arriving at Earth from outer space was first recognized in 1932 by K G Jansky at the Bell Telephone Laboratories in New Jersey He observed a steady arrival of 14 6-m waves from fixed directions in space. The maximum intensity was received from the direction of the center of the Milky Way—our spiral galaxy of stars—and a ridge of strong emission was found along the plane of the Milky Way

Since about 1945 radio studies have been made of the Moon Sun several planets the central region and spiral arms of the Milky Way various types of gaseous nebulae within our galaxy normal spiral and peculiar external galaxies and a large number of unidentified objects A large amount of cosmic radio emission originates in a nearly spherical halo of hot tenuous gas enveloping our Milky Way galaxy and other spiral galaxies Radio waves have been detected from the Sun and it is expected that bursts of radio emission from certain stars will eventually be recorded

Significance of radio astronomy Over most of the electromagnetic spectrum the waves arriving at Earth from interstellar space cannot penetrate Earth s atmosphere and ionosphere Only visible light the nearby infrared and a portion of the radio spectrum reach the surface of Earth. The radio penetration window extends from a short wavelength limit of several millimeters depending upon the state and humidity of the lower atmosphere to a long wavelength limit of tens of meters depending upon the state of the ionosphere

With the advent of artificial Earth satellites and other space vehicles radio astronomy will merge with infrared optical ultraviolet and x ray astronomy thereby connecting all branches of observational astronomy and presenting an unobscured view of the universe except where opaque clouds of interstellar gas or dust may interfere

Radio waves are identical with light waves except that they are much longer This great difference in wavelength results in important and complementary differences between radio and optical astronomy First the radio wavelengths are difficult to focus sharply because wave diffraction requires telescopes many times larger than in the optical case to obtain a corresponding degree of sharpness of detail or angular resolving power (*see* RADIO TELESCOPE) Second radio waves readily penetrate opaque planetary atmospheres and interstellar dust clouds All weather radio observations can be made of celestial regions hidden behind interstellar dust and of planetary surfaces below permanent cloud cover On the other hand radio waves are unable to penetrate the optically transparent ionized gases enveloping the Sun and other stars Long radio waves are absorbed by tenuous transparent, but nearly invisible ionized gases in interstellar space

Several modes of radio observation are common mapping radio intensities over the celestial sphere cataloging position size intensity and polarization of localized intense sources of radio emission and continuously recording transient variations and outbursts of emission from the Sun or the planet Jupiter as a function of time and frequency

When the galactic signal level is higher than the noise generated within the receiver it is possible to listen to the steady hissing noise from the galaxy its characteristics are similar to the noise generated

Frequency shift transmission The carrier frequency is shifted a fixed amount to correspond with telegraphic dots and dashes, or with combinations of pulse signals identified with the characters on a typewriter This technique is widely used in handling the large volume of public message traffic on long circuits, principally by the use of teletypewriters. *See* FREQUENCY MODULATION RADIO

Amplitude modulation The amplitude of the carrier is made to fluctuate corresponding to the fluctuations of a sound wave, a television picture signal, etc This technique is used in AM broadcasting, television picture transmission, and many other services. *See* AMPLITUDE MODULATION

Frequency modulation The frequency of the carrier is made to fluctuate around an average axis to correspond to the modulating wave This technique is used in FM broadcasting, television sound transmission, and microwave relaying. A related technique, called phase modulation, is used in public safety, land transportation, and many other services above 30 Mc. *See* FREQUENCY MODULATION, PHASE MODULATION

Pulse transmission The carrier is transmitted in short pulses, which change in repetition rate, width, or amplitude, or in complex groups of pulses which vary from group to succeeding group in accordance with the message information. These forms of pulse transmission are identified as pulse-code, pulse-time, pulse position, pulse-amplitude, pulse-width or pulse-frequency modulation. Such techniques are complex and are employed principally in microwave relay systems *See* PULSE MODULATION

Radar The carrier is normally transmitted as short pulses in a narrow beam, similar to a searchlight. When a wave pulse strikes an object, such as an aircraft, energy is reflected back to the station, which measures the round trip time and converts it to distance. An airborne radar can receive varying reflections from the earth beneath it and display these reflections in a maplike presentation on a cathode-ray tube. *See* AIRBORNE RADAR, RADAR.

Uses of radio. The first practical application of radio, over a half-century ago between ships and shore stations, was followed quickly by communication overseas and between other widely separated fixed points. Subsequently the applications have become widely diversified. Some of the important uses are listed below

1 Public safety Marine and aviation communications, police and fire protection, forestry conservation, highway traffic control.

2 Industrial Power utilities, pipelines, relay services, news systems, agriculture, petroleum processing.

3 Land transportation Railroads, motor carriers, taxicabs, automobile emergency needs.

4 Broadcasting Television, FM broadcasting, standard broadcasting, short wave international broadcasting.

5 Military Radar, communications, navigation, telemetering, missile tracking, guidance, detection, and aiming.

6 Fixed point-to-point Long-distance message and picture transmission.

7 Relaying Television, sound, picture, and public message relaying over long distances.

8 Telemetering Remote indication of water levels in reservoirs and rivers, performance of experimental aircraft, missile and satellite performance, and other data.

Radio channels. Hundreds of thousands of radio transmitters exist, each requiring a carrier of some frequency If all operated on the same frequency, the interference would be intolerable. This is prevented by using different carrier frequencies for transmitters where service areas overlap, and building receivers which select only the carrier frequency of the desired station. Resonant electric circuits in the receiver are adjusted, or tuned, to accept one frequency and reject others.

Each station operates within a specific rf channel. All other stations within a geographical area are excluded from using this channel. Each channel must be wide enough to accommodate the message information, provide tolerance for small carrier frequency drift, perhaps provide a guard band and allow for imperfect receiver selectivity capabilities. The minimum usable channel widths (or bandwidths) vary from service to service, depending upon the amount of information a channel must accommodate. In television it is 6000 kc, because of the large amount of essential picture information. In FM broadcasting it is 200 kc, and in AM broadcasting 10 kc The great demand for authorizations requires efficient channel utilization. In the mobile transportation service, the FCC was compelled to reduce channel widths 50% when technical developments made it feasible. *See* BANDWIDTH REQUIREMENTS (COMMUNICATIONS)

FCC regulations All nations have a sovereign right to use freely any or all parts of the radio spectrum. But a growing list of international agreements and treaties divides the spectrum and specifies sharing among nations for their mutual benefit and protection.

Every nation designates its own regulatory agency Functioning within the international agreements, it issues authorizations, assigns frequencies, polices operations, creates technical standards, rules and practices, and safeguards and protects the public interest.

In the United States all nongovernmental radio communications are regulated by the Federal Communications Commission according to the provision of the Communications Act of 1934 as amended. Creation of a radio station or service requires authorization by the FCC. Upon completion of an authorized facility a license to operate is issued Radio stations are inspected regularly by engineers attached to FCC field offices. Stations must comply with the terms of their authorization regarding carrier frequency tolerance, power limitations, permissible communications, call signals, and control by properly licensed personnel

Volume requirements The demand for FCC radio authorizations is incr eat fo

fact was suggested earlier by infrared measurements of the Moon

Echoes from the Moon Radar echoes have been obtained from the Moon and from the planet Venus The time it takes a radio pulse to travel from the Earth to the object and back is a measure of the distance to the object From these studies it is possible to deduce a more refined value for the scale of the solar system than has previously been possible The recording of the intensity of radar echoes at different frequencies provides information about the electrical properties of the surface of the Moon or planets, or the atmospheric or interplanetary gases

Centimeter waves from planets Radio waves have been measured from the planets Venus Mars Jupiter ... These waves may be generated thermall ... ense at ... entime ... erature ... f at le ... greater than anticipa ... ations Similar measurements of Jupiter require a temperature of about 150°K near a wavelength of 3 cm and higher variable temperatures at longer wavelengths The emission from Mars is not in ... thermal emission from its solid sur ...

possible to deduce the period of rotation o ... planet Venus by searching for small variations in the intensity of the emitted radio signals

Decameter waves from Jupiter At the radio wavelengths greater than 10 m it has been discovered that Jupiter frequently emits intense narrow band bursts of radio energy A study covering about a decade has disclosed localized regions of radio emission which indicate a fixed statistical relationship in the longitudinal spacing on Jupiter between these active regions This indicates that the source of these radio waves is fixed to the solid rotating body of Jupiter The period obtained is the true period of rotation of Jupiter, it is slightly shorter than the observed periods of rotation of the opaque atmosphere The measurement of polarization of radio waves from Jupiter discloses that an appreciable fraction of the radio waves are circularly polarized This suggests the existence of a magnetic field on Jupiter of about 5 oersteds, or about 10 times the Earth's field See JUPITER

Cosmic background radiation Radio waves are emitted from all directions in space The background radiation intensity increases steadily with increasing wavelengths, and is of both galactic and extr... The extragalactic component is ... we ... as ... galac... disk distribution coinciding with the disk ... Milky Way and slightly flattened spheroidal distribution of radiation enveloping the disk radiation (Fig 2) The spheroidal component has also been observed around a neighboring spiral galaxy in the constellation of Andromeda and is referred to as the galactic corona or halo It radiates by the acceleration of energetic electrons in a magnetic field of the order of 10^{-6} oersted which forms the halo

The galactic disk radiation can be further resolved into at least two components a thermal component due to the emission from interstellar ionized hydrogen and a nonthermal component, which may be due to the energetic electrons either in magnetic fields distributed either continuously throughout the galactic disk or spiral arms or concentrated in discrete localized regions of radio emission The radio spectra of the above components of galactic radio emission are required in the attempt to determine the structure, composition, and development of spiral galaxies

Cosmic radio sources Many localized regions or discrete intense sources of radio emission have

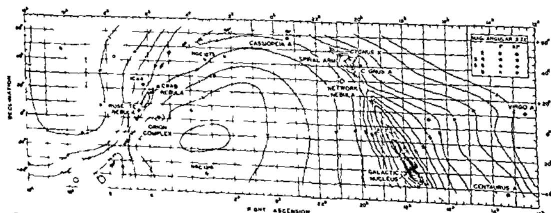


Fig 2 A radio map of the sky at a wavelength of 1 2 m. The contours show the radio brightness of the sky while the small circles indicate discrete radio sources. The solid dots indicate sources less than 1° in diameter the open circles are approximately the angular extent of larger radio sources The celestial coordinates are for epoch 1950 0 (Ohio State University Radio Observatory)

in a receiver Measurements can be made of noise signals some 10,000–100,000 times weaker than the noise level intrinsic to the measuring receiver by integrating the output of a wide band pass receiver for several minutes

Generation of radio waves by nature. The basic mechanism for the generation of radio waves is the oscillation or acceleration of electric charge, principally the acceleration of electrons in electric or magnetic fields, lightning discharges, explo sions, *shock waves in the solar atmosphere, the* acceleration of energetic electrons in magnetic fields (the resulting radiation being called synchro tron radiation), and collisions between electrons and ions or molecules in ionized gas clouds are examples in nature Radio waves are also emitted during electronic transitions between closely spaced atomic and molecular energy levels, mo lecular transitions in the molecules of oxygen and of water vapor in Earth's atmosphere and the atomic transition in the ground state of neutral atomic hydrogen in interstellar gas clouds have been measured *See* ATOMIC STRUCTURE AND SPEC TRA, ELECTROMAGNETIC RADIATION, MOLECULAR STRUCTURE AND SPECTRA

Waves from the Sun's atmosphere Radio waves are emitted from the Sun's atmosphere at a steady level when it is undisturbed and in bursts when disturbed by sunspot or flare activity (Fig 1) The steady level of emission is due to thermal collisions between electrons and ions in the hot tenuous solar atmosphere The long radio waves are emitted from the outer atmosphere or corona Centimeter wave measurements give information about the tempera ture and electron density distribution in the lower solar atmosphere, or chromosphere Large sun spots commonly produce excess radio emission which is closely correlated with the electron con centration in the E layer of the ionosphere *See* IONOSPHERE

Occasionally, active sunspot regions flare up in the spectral lines of hydrogen, helium, and calcium in a localized region on the solar surface. Some flares produce sufficient emission of ultraviolet, x ray, optical line radiation, and of high velocity charged particles to disrupt Earth's ionosphere and prevent world wide radio communications. A day or so later, during the arrival of the slower charged particles, the ionosphere is again disturbed and may remain so for many hours. During these periods, *Earth's magnetic field is affected,* currents are induced in the Earth which disrupt communica tion lines, and auroral displays, or the northern lights, are frequently seen *See* AURORA

Observation of the time variation of the radio spectra of bursts from the Sun, provides means to deduce the velocity at which the charged particles are expelled from the active regions on the Sun. During intense flares, cosmic ray increases have been recorded on Earth, and solar radio spectra observations indicate the temporary existence of energetic electrons that radiate by moving in strong magnetic fields in regions comparable in size to the Sun and which extend outward from the Sun. It is believed that the energetic electrons and the solar cosmic ray particles are accelerated simultaneously by the same mechanism in the solar atmosphere. A corresponding close relationship is believed to exist between galactic cosmic rays and radio emission. *See* COSMIC RAYS

Waves from the Moon Radio waves have been observed from the Moon at centimeter wavelengths. These waves are thermally generated in the outer few centimeters or meters of the Moon's surface layers Measurements of the variation of the radio intensity with frequency and with phases of the moon yield information about the thermal and elec trical properties of the Moon's surface. The evi dence requires the presence of an exceptionally good heat insulating layer of dust on the moon. The

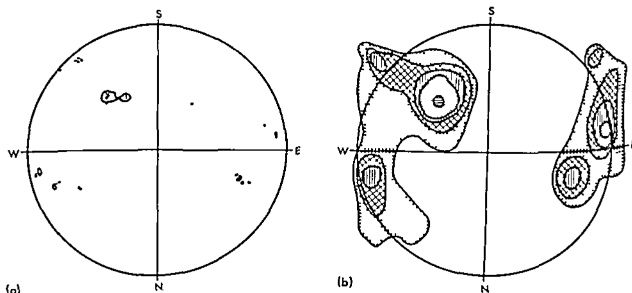


Fig 1 Sun on June 27, 1957 (a) White-light sketch of the Sun showing sunspots (b) Radio picture of the Sun illustrated by contours of the radio brightness The fainter radio brightness contours of the quiet Sun are not shown (*Radiophysics Laboratory, CSIRO, Sydney Australia*)

act was suggested earlier by infrared measurements of the Moon

Echoes from the Moon Radar echoes have been obtained from the Moon and from the planet Venus The time it takes a radio pulse to travel from the ... to the object and back is a measure of the ...nce to the object From these studies, it is ...ble to deduce a more refined value for the ... of the solar system than has previously been ...ible The recording of the intensity of radar ...es at different frequencies provides informa... about the electrical properties of the surface ...he Moon or planets or the atmospheric or inter...etary gases

Centimeter waves from planets Radio waves have ...n measured from the planets Venus, Mars, ...piter and Saturn These waves may be generated ...rmally in the solid or liquid surface layers or ... atmospheres of the planets The intensity of

...ations Similar measurements of Jupiter require a ...emperature of about 150°K near a wavelength of ... cm and higher variable temperatures at longer

possible to deduce the period of rotation of the planet Venus by searching for small variations in the intensity of the emitted radio signals

Decameter waves from Jupiter At the radio wavelengths greater than 10 m it has been discovered that Jupiter frequently emits intense narrow band bursts of radio energy A study covering about a decade has disclosed localized regions of radio emission which indicate a fixed statistical relationship in the longitudinal spacing on Jupiter between these active regions This indicates that

th...

rotati...

true period of rotation of Jupiter, ... shorter than the observed periods of rotation of the opaque atmosphere The measurement of polarization of radio waves from Jupiter discloses that an appreciable fraction of the radio waves are circularly polarized This suggests the existence of a magnetic field on Jupiter of about 5 oersteds, or about 10 times the Earth's field See JUPITER

Cosmic background radiation. Radio waves are emitted from all directions in space The background radiation intensity increases steadily with increasing wavelengths, and is of both galactic and extragalactic origin The extragalactic component is probably the result of unresolved and numerous weak extragalactic sources of radio emission, such as normal, peculiar, or colliding galaxies The galactic component can be resolved into a flattened-disk distribution coinciding with the disk of the M... ...and slightly flattened spheroidal dis...

tr...

(...

ob...

constellation of Andromeda and is referred to as the galactic corona or halo It radiates by the acceleration of energetic electrons in a magnetic field of the order of 10^{-6} oersted which forms the halo

The galactic disk radiation can be further resolved into at least two components, a thermal component due to the emission from interstellar ionized hydrogen and a nonthermal component, which may be due to the energetic electrons either in magnetic fields distributed either continuously throughout the galactic disk or spiral arms or concentrated in discrete localized regions of radio emission The radio spectra of the above components of galactic radio emission are required in the attempt to determine the structure, composition,

... regions
...ion have

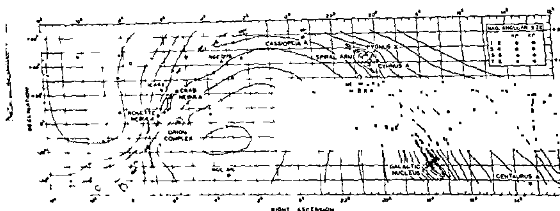


Fig 2 A *radio map of the sky at a wavelength of* 1.2 *m. The contours show the radio brightness of the sky while the small circles indicate discrete radio sources. The solid dots indicate sources less than* 1° *in diameter, the open circles are approximately* ... *angular extent of larger radio sources The celes... coordinates are for epoch* 1950.0 (*Ohio State Univ... sity Radio Observatory*)

been found distributed over the celestial sphere Certain galaxies external to our system emit radio waves both in the hydrogen line and as continuous spectrum radiation The hydrogen line radiation has been detected from the Magellanic Clouds, two nearby external galaxies Continuous radio emission has been observed from a number of known galaxies, some are normal spiral galaxies similar to the Milky Way and radiating at about the same radio energy rate, the others are peculiar galaxies, mostly pairs of colliding galaxies, radiating at a rate several orders of magnitude greater than the normal spirals It is believed that a number of the known radio sources are external galaxies, perhaps pairs of colliding galaxies beyond the range of optical telescopes

Only a small fraction of the radio sources so far detected has been identified with optical objects There are at least two spectral types of radio sources a type having a radio emission spectrum consistent with thermal radiation from ionized interstellar gas, and a nonthermal type whose radio spectrum is not in accordance either in form or in intensity with thermal radiation but is consistent with radiation from the acceleration of energetic electrons in interstellar magnetic fields Because of their different spectra, the nonthermal sources are prevalent at meter wavelengths and the thermal sources at centimeter wavelengths The thermal sources are bright galactic nebulae, such as the Great Nebula in Orion, the nonthermal sources include normal spiral galaxies like the Andromeda Nebula, colliding galaxies like the intense source in Cygnus, unusual galactic nebulosities like the intense source in Cassiopeia, and supernovae remnants like the Crab Nebula in Taurus (*see* CRAB NEBULA) This is the first cosmic radio source to be identified with a celestial object It is the result of a rare explosion of a star in the year 1054 A D , the gases are still expanding at a velocity of 1100 km/sec The light from the central amorphous mass within the filamentary shell is linearly polarized to such a degree that its explanation requires the assumption that this light is generated by the acceleration of energetic (10^{12} ev) electrons in a magnetic field (about 10^{-4} oersteds) Centimeter waves emitted by the Crab Nebula are also partially polarized These facts and its radio spectrum, strongly suggest a similar mechanism for the origin of its radio emission

The distribution of thermal sources is concentrated toward the galactic equator, the plane of the Milky Way This is consistent with the distribution of bright galactic nebulae which occur in the spiral arms Nonthermal sources are distributed in two classes Class I consists of large intense sources which are concentrated near the galactic equator and have a measured angular size of about 1° Class II sources show little concentration toward the plane of the galaxy and are usually a few minutes of arc or less Class I sources contribute to the continuous background emission concentrated along the galactic equator Class II sources consist of both galactic and extragalactic sources

There is one radio source that is in a class by itself It is an intense source in the galactic nucleus, some galaxies similar to the Milky Way have an intense concentration of stars in the center of the bright central bulge This dense nuclear concentration of stars may also exist in our galaxy hidden behind concentrations of interstellar dust, but disclosing its presence by intense radio emission

Radio sources are of intrinsic value They generate radio energy by a process that is also at work in the solar atmosphere They supply clues to the processes of galactic evolution and to the type of cosmology governing the universe

Hydrogen line radiation. Radio waves are emitted and absorbed by neutral atomic hydrogen gas at a wavelength of 21 106 cm This corresponds to a transition between two closely spaced energy levels in the hyperfine structure of the ground state. The majority of the gas in the galaxy is hydrogen the bulk of it is neutral and atomic It occurs in clouds concentrated in the galactic plane, principally along the spiral arms Measurement of the line emission profiles in the plane of the galaxy gives the Doppler shift in frequency, which yields a measure of the relative radial velocity between Earth and interstellar clouds within the antenna beam The combination of these velocities determined along the galactic equator with a model of galactic rotation found by optical studies makes possible the mapping of the spiral form of the gaseous arms Investigations of this type have revealed for the first time the nearly complete spiral pattern of the Milky Way galaxy This is an outstanding accomplishment of radio astronomy Opti-

uring the Doppler velocity and radio intensities surrounding localized interstellar regions in an attempt to obtain data on the physical processes involved in the formation of new stars

By the measurement of the exact frequency of the 21 cm absorption lines due to interstellar clouds in front of distant radio sources emitting a continuous spectrum it is possible to obtain information on the structure and size of these clouds of neutral hydrogen It may be possible to detect a Zeeman splitting of the lines due to weak galactic magnetic fields and thereby measure these fields. *See* ZEEMAN EFFECT

There is a possibility of detecting the emission or absorption of other radio spectral lines, a neutral atomic deuterium line occurs at a wavelength of 91 57 cm, there are two lines of the hydroxyl radical at 17 98 and 18 00-cm wavelength, and several lines of hydrogen and helium at other wavelengths

Cosmology. Because the second brightest radio source is believed to be about 500,000 000 light years distant and because it would be possible to detect this source if it were ten or more times further away, it has been ass at a large umber

f the unidentified radio sources are too distant to e photographed by the 200-in Hale telescope at Iount Palomar For this reason, radio telescopes re believed to be penetrating deeper into space han optical telescopes and thus will play an im portant role in determining a cosmological model appropriate to the universe

Various types of cosmological models can be selected or rejected by counting the number of radio sources in each intensity interval or by measuring the radio source size distribution of many faint sources

Observations from satellites Use of small, low gain antennas and a radio receiver in a space vehicle extends the radio energy spectrum of the [illegible] background radiation from a wavelength

[illegible]

spectra of solar bursts and perhaps bursts [illegible] Jupiter can also be extended to longer wavelengths by means of a low gain antenna and a sweep-frequency receiver in an artificial satellite orbiting high above Earth s ionosphere

[illegible]

the galactic background intensity at low frequencies, to measure the low frequency spectrum of the undisturbed Sun and to make studies of low fre[qu]ency bursts from Jupiter and possibly from [ot]her planets and stars Space probes which pene[tr]ate below the ionospheres on Venus and Jupiter [a]nd measure radio waves that may be trapped be[lo]w the planetary ionospheres would be of interest [W]ith the advent of a radio observatory on the [M]oon it will be possible to observe the entire radio [s]pectrum from a wavelength of tens of kilometers [t]o the infrared region Large antennas can be [e]rected free of wind and weather in the reduced gravitational field of the Moon

Applications Radio waves from the Sun Moon and radio sources can be used for all weather *navigation* because radio waves penetrate rain fog, snow and clouds The accuracy of a solar sextant developed by the U S Navy exceeds that obtainable by the optical hand sextant Large radio antennas can be calibrated by using noise signals from the bright radio sources planets the Moon and the Sun, which have been measured by radio-astronomers Radio sources are used to study radio propagation through the Earth's lower atmosphere and ionosphere. Cosmic radio noise levels are required in the design of radio communications, navigation and detection systems Large radio antennas designed for radio astronomy are now used for radar radio communications and satellite telemetering Special radio techniques have been developed which are now used in tracking artificial Earth satellites and space vehicles. *See* ANTENNA (AERIAL), ASTRONOMICAL SPECTROSCOPY MICROWAVE RADIO-WAVE PROPAGATION [F T H]

Bibliography R D Davies and H P Palmer, *Radio Studies of the Universe* 1959, F T Haddock ([illegible]) Proc IRE Radio Astronomy Issue, 46(1), [illegible]

Radio broadcasting

The transmission of radio programs intended for public reception

Radio broadcasting began in 1920 at Pittsburgh, based upon experimental broadcasting over an amateur radio telephone station owned and operated by Dr Frank Conrad of Westinghouse The conception of a public broadcasting service arose from the great interest shown in these experimental broadcasts of speech and music and led to the es[tab]lishment of station KDKA, the world's first sta

[illegible]

history By the end of one year the number of stations had increased to 31 and at the end of two years over 600 stations were providing service throughout the United States, with growth accelerating rapidly

Scope Radio broadcasting is a highly dependable and extremely valuable service in all civilized areas of the world In many areas of the United States listeners have a choice of as many as 30 stations Over 160 000 000 broadcast receivers are in use in the United States Service is provided for entertainment, education, information and alleviation of distress from early morning until late at night, and in many areas it continues around the clock to serve all segments of the population About 50 000 persons are employed directly in the operation of stations and many more are engaged in allied services, such as apparatus manufacturing and sales script writing and program production, advertising agency activities news and publicity, publication of magazines and trade publications, common carrier network services, and federal regulation

Frequency allocations. By international agreement, certain portions of the radio-frequency spectrum are set aside for radio broadcasting In the United States over 3200 stations occupy the Standard Broadcast Band of 107 channels, extending from 535 to 1605 kilocycles (kc), each channel being 10 kc wide In addition, there are 540 stations on 100 channels utilizing the frequency modulation band from 88 to 108 megacycles (Mc) each band being 200 kc wide Television broadcasting occupies bands in the regions 54–72, 76–88, 174–216, and 470–890 Mc, each channel being 6 Mc wide *See* AMPLITUDE MODULATION RADIO, FREQUENCY MODULATION RADIO, TELEVISION For supplementary and auxiliary service, such as relaying programs from a remote point of origin, transmitting the program from the studio to the transmitter conveying program production orders and communications, and for emergency use frequencies in the areas of 1620 kc, 26 Mc, 152 Mc, 450 Mc and 950 Mc are utilized in the United States *See* RADIO SPECTRUM ALLOCATIONS.

been found distributed over the celestial sphere Certain galaxies external to our system emit radio waves both in the hydrogen line and as continuous spectrum radiation The hydrogen line radiation has been detected from the Magellanic Clouds, two nearby external galaxies Continuous radio emission has been observed from a number of known galaxies, some are normal spiral galaxies similar to the Milky Way and radiating at about the same radio energy rate, the others are peculiar galaxies, mostly pairs of colliding galaxies, radiating at a rate several orders of magnitude greater than the normal spirals It is believed that a number of the known radio sources are external galaxies, perhaps pairs of colliding galaxies beyond the range of optical telescopes

Only a small fraction of the radio sources so far detected has been identified with optical objects There are at least two spectral types of radio sources a type having a radio emission spectrum consistent with thermal radiation from ionized interstellar gas and a nonthermal type whose radio spectrum is not in accordance either in form or in intensity with thermal radiation but is consistent with radiation from the acceleration of energetic electrons in interstellar magnetic fields Because of their different spectra, the nonthermal sources are prevalent at meter wavelengths and the thermal sources at centimeter wavelengths The thermal sources are bright galactic nebulae such as the Great Nebula in Orion, the nonthermal sources include normal spiral galaxies like the Andromeda Nebula, colliding galaxies like the intense source in Cygnus unusual galactic nebulosities like the intense source in Cassiopeia and supernovae remnants like the Crab Nebula in Taurus (*see* CRAB NEBULA) This is the first cosmic radio source to be identified with a celestial object It is the result of a rare explosion of a star in the year 1054 A D, the gases are still expanding at a velocity of 1100 km/sec The light from the central amorphous mass within the filamentary shell is linearly polarized to such a degree that its explanation requires the assumption that this light is generated by the acceleration of energetic (10^{12} ev) electrons in a magnetic field (about 10^{-4} oersteds) Centimeter waves emitted by the Crab Nebula are also partially polarized These facts and its radio spectrum strongly suggest a similar mechanism for the origin of its radio emission

The distribution of thermal sources is concentrated toward the galactic equator the plane of the Milky Way This is consistent with the distribution of bright galactic nebulae which occur in the spiral arms Nonthermal sources are distributed in two classes Class I consists of large intense sources which are concentrated near the galactic equator and have a measured angular size of about 1° Class II sources show little concentration toward the plane of the galaxy and are usually a few minutes of arc or less Class I sources contribute to the continuous background emission concentrated along the galactic equator Class II sources consist of both galactic and extragalactic sources

There is one radio source that is in a class by itself It is an intense source in the galactic nucleus, some galaxies similar to the Milky Way have an intense concentration of stars in the center of the bright central bulge. This dense nuclear concentration of stars may also exist in our galaxy hidden behind concentrations of interstellar dust, but disclosing its presence by intense radio emission

Radio sources are of intrinsic value They generate radio energy by a process that is also at work in the solar atmosphere They supply clues to the processes of galactic evolution and to the type of cosmology governing the universe

Hydrogen line radiation. Radio waves are emitted and absorbed by neutral atomic hydrogen gas at a wavelength of 21 106 cm This corresponds to a transition between two closely spaced energy levels in the hyperfine structure of the ground state. The majority of the gas in the galaxy is hydrogen the bulk of it is neutral and atomic. It occurs in clouds concentrated in the galactic plane principally along the spiral arms Measurement of the line emission profiles in the plane of the galaxy gives the Doppler shift in frequency, which yield a measure of the relative radial velocity between Earth and interstellar clouds within the antenna beam The combination of these velocities determined along the galactic equator with a model of galactic rotation found by optical studies makes possible the mapping of the spiral form of the gaseous arms Investigations of this type have revealed for the first time the nearly complete spiral

uring the Doppler velocity and radio intensities surrounding localized interstellar regions in an attempt to obtain data on the physical processes involved in the formation of new stars

By the measurement of the exact frequency of the 21 cm absorption lines due to interstellar clouds in front of distant radio sources emitting a continuous spectrum it is possible to obtain information on the structure and size of these clouds of neutral hydrogen It may be possible to detect a Zeeman splitting of the lines due to weak galactic magnetic fields and thereby measure these fields *See* ZEEMAN EFFECT

There is a possibility of detecting the emission or absorption of other radio spectral lines, a neutral atomic deuterium line occurs at a wavelength of 91 57 cm there are two lines of the hydroxyl radical at 17 98 and 18 00-cm wavelength and several lines of hydrogen and helium at other wavelengths

Cosmology Because the second brightest radio source is believed to be about 500 000 000 light years distant and because it would be possible to detect this source if it were ten or more times farther away it has been assumed that a large number

other stations having higher priority status on the channel

Clear channel A channel on which the dominant station or st ... de areas is called a c ... are cleared o ... ir primary service areas and over all or a sub- ntial portion of their secondary areas

\ lear channel which has no duplication at ...

protected against interference to its 100 ... lt ground wave contour from cochannel stations ... ground wave contour from ...

ant Mexican stations 1 to the Bahamas ... emainder to dominant stations in the United tates

On class IB clear channels night duplication is permitted if interference is not created at those ... where the 500-microvolt nighttime signal ...

other countries

A class II station is a secondary station which operates on a clear channel and which must protect the dominant class I station or stations by limitations in power or operating hours or by use of a directional antenna

Regional channel Several geographically separated stations may operate on a regional channel with powers not in excess of 5 kw The primary nighttime service area of a regional channel station may be substantially limited as a result of interference from other cochannel stations Its purpose is to serve regional areas

The following frequencies are assigned for use by regional stations 550–630 790 910–930 950–980 1150 1250 1330 1350–1390 1410–1440 1460–1480 1590 and 1600 kc

Local channel Many geographically separated stations may operate on a local channel with powers not in excess of 250 watts The primary service area of a local channel station may be severely limited as a consequence of interference from other cochannel stations Its purpose is to serve local areas Six frequencies between 1230 and 1490 kc are assigned for use by local stations

Field intensity measurements Field intensities may be estimated closely by calculation but for accurate determination measurements are necessary They frequently are required for evidence in FCC hearings and are assembled by most stations for a measure of effectiveness and for sales promotion

The conventional method of making a field intensity survey is to lay out on a road map at least eight routes corresponding to equally spaced radials going out from the transmitter The field intensity is measured at intervals of a mile or two beginning at the transmitter and ending at the distance where the field intensity is too low to be of interest

Upon completion of the survey the measurements are plotted on graph paper and a smooth curve is drawn through them This information can be transferred to a map and the points of equal field intensity on the various radials are connected to form smooth contour lines (see Fig 2) For precise measurement of the 1 mile field, measurements at close intervals are made to bracket the 1 mile distance

The useful coverage of a broadcast station is evaluated in terms of the distances or areas over which field intensities of interest are provided When the natural static level and the receiver input noise level are known the field intensity required to override them may be easily derived to determine the minimum values required for satisfactory reception It is common practice to use a value of 0 5 millivolts per meter (mv/m) as the minimum field intensity required for primary service in open country For a large heavily populated suburban area 10 mv/m is often used

Station coverage calculations In planning the construction of a station and particularly its location the coverage areas are estimated closely in advance Three important factors are the power, the antenna efficiency and the field intensity which they will produce at some short distance where earth attenuation is negligible The distance used is ordinarily 1 mile and the 1 mile field intensity is a commonly used figure of merit For an efficient

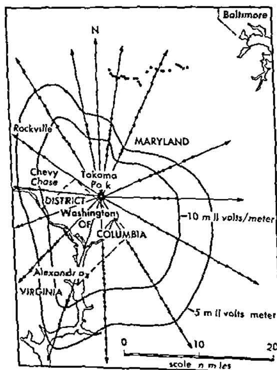


Fig 2 Example of a field intensity survey

FCC requirements. All radio stations of any kind in the United States must be licensed by the Federal Communications Commission (FCC) In response to a written application to construct a broadcast station in accordance with specified conditions, FCC issues a construction permit provided the station will not create prohibitive interference for existing stations, and provided the station will serve public interest, convenience, and necessity When all conditions are met, a license to operate the station is issued Some important technical conditions to be met follow

Frequency tolerance The maximum permissible deviation of the assigned carrier frequency from its specified value is limited for standard broadcasting stations to plus or minus 20 cycles This limitation minimizes interfering beat note audibility

This close tolerance is met by the use of crystal controlled vacuum tube oscillators, in which special precautions are taken to minimize the effects of changes in ambient temperature and circuit constants With modern equipment, variations are limited to only a few cycles per second over long periods of time

Carrier power The power of the radio frequency carrier is specified in the FCC construction permit and license Licensed carrier powers range from 100 watts for small stations serving local areas to 50,000 watts for the largest stations serving large areas and long distances The specified powers must be rigidly maintained

Directional antenna usage Often it is possible to avoid interfering with another station by the use of a directional antenna which radiates low power in the direction of another station assigned to the same channel and concentrates its coverage in directions in which lie the areas most desirable to cover Figure 1 shows a typical directional antenna pattern

Signal propagation Signals from standard broadcast stations arrive at the receiving points in two ways, by the ground wave and by the sky, or ionospheric, wave In the daytime only the ground wave is received but at night both ground and sky waves are effective See RADIO WAVE PROPAGATION

The variation of radio field intensity caused by changes in wave propagation in the ionosphere is known as fading From early evening until dawn radio waves at standard broadcast frequencies are reflected back to earth from the unstable E layer of ionized air about 60 miles above the earth Changes in reflection may cause fluctuations which vary widely with time These reflected sky waves return to the earth at distances from 50 to 1000 miles However, propagation may occur over thousands of miles from multiple reflections between earth and ionosphere Deep fading may be present when sky waves and ground waves are present simultaneously and have comparable amplitudes but opposite phase

Broadcast service area The geographical area over which a standard broadcast station projects its signals is divided by FCC into two categories

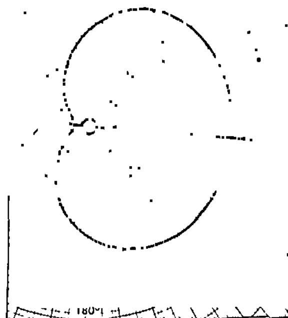


Fig 1 Example of a directional broadcast antenna pattern

A primary service area is the area served by ground-wave propagation with sufficient field intensity to provide consistent day and night service. A secondary service area is the area served by ground wave

conducting, the signal strength would fall off directly with distance but the actual decrease in signal is greater, because the earth is not a perfect conductor

Signal strength at a receiving point is termed its field intensity which is commonly expressed in terms of the root mean square value of the electric field in the direction of maximum intensity Because a greater voltage will be produced by a more effective receiving antenna the field intensity is expressed as so many microvolts per meter, the latter portion of the term indicating the effectiveness of the receiving antenna

Broadcast channel classification. Except for a limited number of channels reserved for service over large areas each broadcast channel is used by a large number of geographically separated stations All these stations do not have the same priority status A dominant station may be permitted to broadcast full time Some secondary stations are permitted to operate only between the hours of sunrise and sunset so they will not interfere with

ey arrive at the reflecting ionized layers of air lled the F1 and F2 layers where they are re ected back to earth A number of succeeding re ections between the F layers and the earth make ssible transmission for many thousands of miles d at times transmission completely around the orld back to the starting point

By World War II short wave international roadcasting was conducted by most nations in the orld Such operation is predominantly but not xclusively conducted by the governments of the e pective countries or by organizations controlled y them Commercial operation of short wave roadca ting was authorized and conducted in the Jn ted States by several organizations prior to World War II but such operation plays a minor art at present *See* RADIO RADIO BROADCASTING NETWORKS TELEVISION NETWORKS [R F CU]

Bibliography W L. Everitt *Fundamentals of Radio and Electronics* 1958 Federal Communica tions Comm s on *R iles and Regulations* pt 3 K Henn 5th ed neering ers, *Engineering Handbook*

Radio broadcasting networks

A group of broadcast stations connected by radio or wire so that all stations can broadcast the same p ogram s m ltaneously

a service by which elaborate and expensive pro grams and events of widespread interest are made ava lable to networks of stations all of which share the burdensome costs These events led to the for mation of private network broadcasting companies wh ch contract with AT&T for distribution of their programs

United States broadcasting networks In the United States there are four principal sound broad casting networks and three for television all of whi h ut lize teleph ne company facilities for pro gram distribution The AM networks are listed below with approximate statistics During the sum mer the circu t mileage may increase because of the large dayl ght saving time differentials.

Network	Stations	Circu t miles
Nat onal Br adcast ng Co Inc	200	17 500
Columbia Broa lcast ng System	200	16,500
American Br adcasting Company	284	21 000
[illegible] :m	110	28,000

The television networks are those of NBC, CBS and ABC, and are roughly comparable in size to the AM networks of these companies *See* TELEVISION NETWORKS

+ live program distribution net

of modern broadcasting program dis works take several forms all of which may be an integral part of a national network

Round robin A network circuit that extends from a network office to and through other communities on a circuitous route back to the starting point serving individual stations and groups of stations en route is called a round robin

This type of network operation has the advantage that programs may be introduced from any office on the network by opening the round robin at one of fice and closing it at all other offices The illustra tion shows how a round robin might be constituted The round robin connects the network offices in New York Chicago and Washington Along the route there are Bell system offices, which feed local stations or groups of stations Round robin net works are used for both radio and television broad casting

Round robins simplify the switching of programs The number of stations connected to a network at any one time may change widely throughout the day Stations take network programs at certain times and at other times originate their own

A program originating in New York may be ro ted from the network headquarters to the New

point, if desired the round robin may be opened and a different program sent on through the Bell offices in Indianapolis South Bend and Pittsburgh to an additional group of stations Thus any de sired number of stations may be served by this round robin either individually in groups or in to tal An important feature of a round robin is that the network may perform the switching functions in its own offices without dependence upon remote Bell system offices

The circuit arrangement associated with the Pittsburgh office is shown in some detail to illus trate an intermediate Bell system office At each of the AT&T offices often referred to as repeater sta tions, one or more broadcast stations may be con nected Separate circuits transmit the program from the Bell office to the local broadcast studio After monitoring the program is fed to the local transmitter on another circuit At intervals along the round robin network extensions may be con nected to serve other parts of the country not cov ered by this round robin

The round robin type of circuit provides for transmission in one direction only so that to reach from one office to another substantially the whole round robin may be required For example a trans mission from New York to Washington requires

50 kw nondirectional station the 1 mile field is usually about 1700 mv/m, for 5 kw with a less elaborate antenna it is about 500 mv/m, and for 1 kw with a small antenna it is about 180 mv/m

Another important factor is the attenuation produced by heating losses in the earth which are cumulative and increasingly severe at distances beyond 1 mile These losses vary widely for various types of terrain with salt water contributing little and sand and rock contributing heavily Also the losses for a given type of terrain increase rapidly with the carrier frequency For example a 50 kw 500-kc station with 1700 mv/m at 1 mile transmitting over moist farm land has a primary service area over 10 times as large as a 50 kw 1500 kc station transmitting over a rocky or sandy area

Transmitter locations Transmitter locations are selected with care to provide adequate signal intensity for the maximum number of persons In urban areas interference exists from electrical machinery and high signal attenuation is introduced by the concentrations of large buildings and the frequent location of receivers (with built in antennas) in the lower floors of large hotels and apartment houses To provide noise free service under such conditions transmitters are located in the most favorable direction and usually at such a distance as to provide at least 50 mv/m over the urban area

For suburban areas with more favorable conditions 10–50 mv/m is desirable For rural areas 2–10 mv/m is desirable but the nominal primary service area extends out to the 0 5 mv/m contour Under favorable conditions good service is obtained out to the 0 1 mv/m contour but it is quite vulnerable to static and man made noise Stations of 50 kw are usually located within 15 miles of the urban center 5 kw stations within 5 miles and lower powered stations are usually in the close suburbs or in the urban area itself The FCC discourages locating so close to the heavily populated areas that overpowering signal strength blankets reception of other local stations and it imposes the condition that the number of persons within the 1000 mv/m contour shall not exceed 1% of those within the 25 mv/m contour

Frequency modulation broadcasting Frequency modulation (FM) came early in radio development Patents on FM transmission were requested in 1902 and granted in 1905 By 1935 techniques and devices had been developed and combined in a system which provided unique advantages over regular amplitude modulation (AM) broadcasting Edwin H Armstrong was closely identified with the development and growth of FM broadcasting

FM broadcasting has full provision for high fidelity reproduction is ordinarily free of static substantially suppresses noise and interference and is relatively free of fading The service range is ordinarily 40–80 miles but stations with high power and high antenna altitude are often receivable at distances of 150 miles

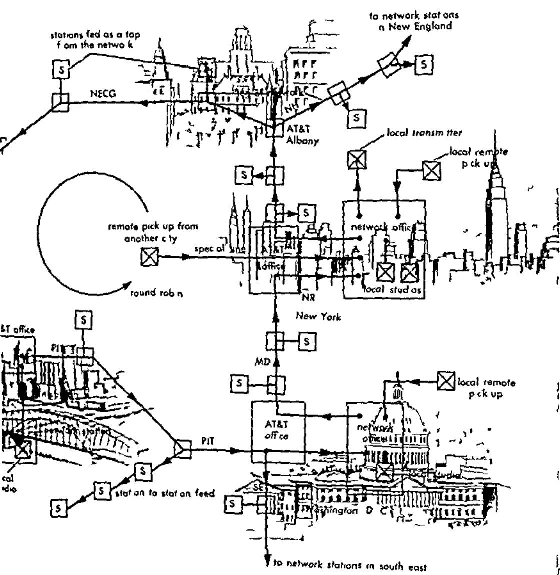
Receivers made exclusively for AM broadcasting cannot receive FM and vice versa but combination receivers are manufactured which can receive both systems

In the United States and many other countries FM broadcasting is assigned frequencies between 88 and 108 Mc. The channel widths are 200 kc. Commercial FM broadcasting was authorized by the FCC on October 31, 1940 In the first year 25 stations were authorized in the United States. This number has increased to 690 The majority of FM stations are operated by owners of AM stations and the stations usually broadcast the same programs, to enable listeners to enjoy FM high fidelity reception of their favorite programs As an economic aid to FM broadcasters the FCC in 1955 authorized supplementary functional music transmission to provide subscribers with background music in restaurants offices stores and factories This is permissible by conventional FM transmission or by the process of multiplexing in which the second program may be simultaneously transmitted and received by the use of a subcarrier Without multiplexing functional music may be broadcast only when the regular FM program is absent With further growth of FM broadcasting and more experience with multiplexing techniques the requirements may be changed to specify that functional music or other auxiliary services must utilize multiplexing to avoid suspension of the regular broadcasting program service

The FCC permits experimental stereophonic broadcasting utilizing two separate but related audio channels The wide FM channel makes it seemingly more feasible to do this in FM than in AM broadcasting which is limited to a bandwidth of only 10 kc

To understand how a stereophonic system operates assume that a symphonic orchestra is being broadcast Without the stereophonic feature, the entire orchestra may be picked up by a single microphone and reproduced on a single loudspeaker with the result that the listener has no sense of the location of the performing artists In the stereophonic system however one microphone would be at the left center of the orchestra and another at the right center At the receiver there would be two correspondingly separated loudspeakers The microphone at the left would provide a signal exclusively for the loudspeaker at the left and the right hand microphone would exclusively excite the right hand speaker thus giving a greater sense of spatial realism and breadth Following further development of stereophonic techniques the FCC may adopt a standard method of FM stereophonic transmission and reception and authorize commercial transmission

Short wave international broadcasting Shortly after World War I experiments in transmission on radio frequencies between about 4000 and 20 000 kc proved that communication was possible over great distances by reflection from layers of ionized air which vary in altitude from about 150 to 250 miles Radio waves from the transmitting antenna follow a path at low angles above the earth until



the stations they feed There are many such legs wh ch feed geographical regions such as the New England states the southeastern states and the n rth central states

To originate a program for the network from a station on a one-way leg various methods are pos ble One is to lease a special circuit from the net work leg station to the nearest AT&T point on the round rob n and utilize the AT&T staff for the switching operation Another is to order the cir cu t from the network leg station to the nearest net work c ntr l off ce which performs the switching operation Connecti n by way of the network office has the advantage that the network can adjust to any sch edule change independently of AT&T

Tap from the network This comprises one indi vidual stati n fed from a network line

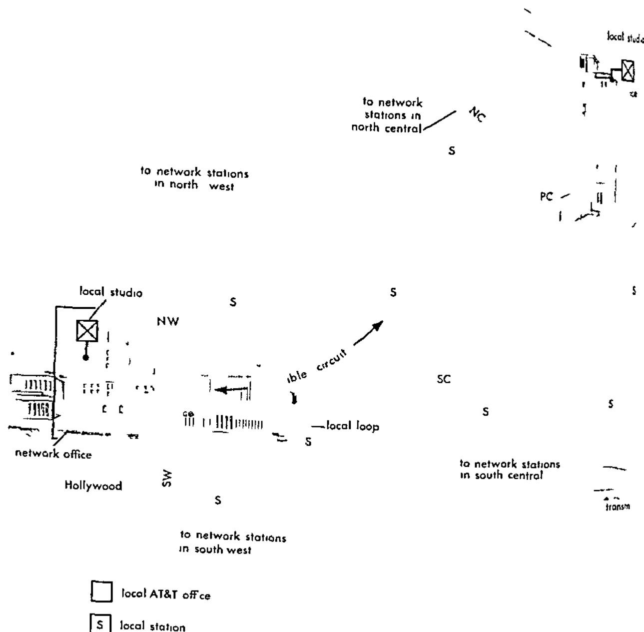
Station to-station feed With this type of network operati n a program is sent to a station which in turn sends it on a Bell circuit to a second station whi h in turn sends it to a third and so on

Local loop The telephone line which connects a telephone-system office to a broadcasting station is called a local loop The loop circuits are owned and operated by local telephone companies

Special lines These circuits are provided usu ally on a temporary basis to send a program over considerable distances such as between cities Such lines are used often for remote pickups from points to which regular lines are not established Network broadcasting stations are not fed from special lines until after the program has gone through the net work control point

In-and out feeds The Bell system circuits at their local offices are connected to the local offices of the network or the individual broadcasting stations by in and-out feeds These feeds make it possible for the customer to perform switching functions be cause he has at his disposal both the incoming and the outgoing circuits.

Circuit designations Each segment or branch of a network has an identifying designation for con



Typical network single line block diagram

that the program go through Albany, Chicago, and Pittsburgh

In practical operation of a round robin two or more network offices may participate in the production of a program For example, it may be desired to have an announcer in New York for a program originating in Chicago With control of the network from the broadcast offices the switch between the program and announcer may be made instantaneously on word cues This is one of the reasons why the round robin was established with the network offices as the nucleus

At the network offices, highly skilled and experienced technical personnel make a great many program switches quickly and simply

Reversible circuit A broadcast network circuit in which the direction of transmission may be reversed at any of a number of intermediate points or terminals is known as a reversible circuit It is possible to have continuous transmission from one end to the other, reverse it in direction or open the circuit at an intermediate office and transmit a pro-

tion is performed by the AT&T staff by prearrangement

Reversal switching requires some loss of program time and a sponsor may be willing to spend thousands of dollars to avoid it even though only *seconds* are involved The reversible circuit lacks some of the flexibility of the round robin circuit It normally is controlled from designated points by electrically operated relays

robin system, is called the leg of a network Network legs are usually, but not necessarily, one-way circuits from the AT&T office to the leg offices and

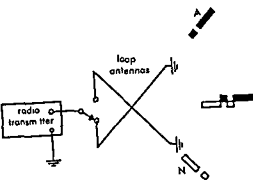
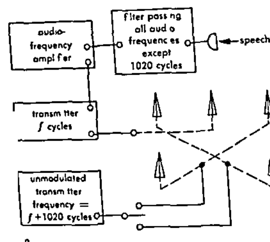


Fig 2 D agrammat cal representation of the princi- ples of the aural rad o range

The *two loops* are connected to the transmitter through a goniometer which permits the rotation of the courses. Provisions are also made to bend and shift the courses in order to align them with the position of the desired airways The width of the course is about 3°

Due to night effect (caused by mixing of the ground and sky wave signals) the loops are often replaced by directional arrays consisting of four vertical radiators (two vertical radiators replacing one loop) This arrangement is known as the T L range To permit simultaneous reception of voice and guidance signals a fifth tower is installed in

keyed ... moved from the guidance-signal transmitter, con sequently, the guidance tone is generated by the beating together of the frequencies from the two



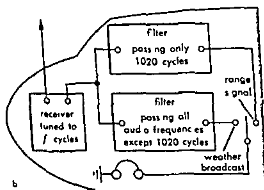


Fig 3 Princ ples of the simultaneous rad o range a) Ground stat on (b) Aircraft

except 1020 cycles Two filters are connected to the aircraft receiver output. One of these filters passes only 1020 cycles (the guidance signal), while the other passes all frequencies except 1020 cycles and is therefore used when voice reception (of weather information) is desired The principle of this simultaneous radio range is shown in Fig 3

If on the other hand the receiver is in line with the plane of the second loop the Morse letter N (dash dot) may be heard. However if the receiver is on a line forming the bisector of the planes of two loops, there will be heard a continuous dash formed by the interlocking of the two letters If an aircraft goes off this course one signal (A or N) is received with greater intensity breaking up the continuous tone and warning the navigator he is off course

Consol. Also called Consolan and Sonne this radio range navigation aid provides a number of characteristic signal zones that rotate in a time sequence A bearing is determined by observation of the instant at which transition occurs from one zone to the following zone

The Consol system consists of three antennas spaced in line. The spacing between the antennas of a pair is approximately 2 88 wavelengths The central antenna of the trio is excited with an un changing radio-frequency current, whereas the other two antennas are supplied with currents that respectively lead and lag by 90° the current in the central antenna Further, the phases of the cur

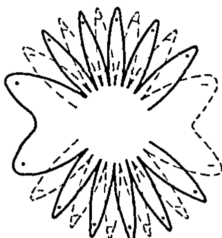


Fig 4 F eld pattern of a Consol system

venience These designations facilitate quick reference to a segment upon which there may be trouble or which may be ordered separately for program connections For example, the NH circuit in the illustration consists of a group of stations in New England fed from the AT&T office in Albany, and IND identifies the section of the round robin circuit between Chicago and Pittsburgh

Channels for program transmission Network channels with different electrical characteristics are available for different purposes The subscriber chooses the type most suitable for his purpose The available channels are listed below, AAA is the most expensive and E the least

Schedule	*Use*	*Frequency range*
AAA	Continuous when so ordered	50–15 000
BBB	Occasional	50–15 000
AA	Continuous, when so ordered	50–8000
BB	Occasional	50–8000
A	Continuous, when so ordered	100–5000
B	Occasional	100–5000
C	Continuous	200–3500
D	Occasional	200–3500
E	Occasional	300–2500, for speech only

Velocity of propagation A radio network may use program lines in which the velocity of propagation is relatively low This is inconsequential with respect to other times involved However, when transmitting a network program of television sight and sound it is necessary to maintain synchronism of lip movements and the velocity of propagation of the sound program lines must closely approximate that of the picture program line Therefore television audio lines must be selected from the network circuits which have low time delay These may consist of carrier circuits rather than metallic circuits To maintain lip synchronism in television the over all delay is limited to about 50 milliseconds compared with the picture

Delay equalization On metallic circuits the low frequency waves normally travel more rapidly than high frequency waves because of circuit reactance This results in arrival of low frequencies before high frequencies and production of distortion of the sounds which would be objectionable To avoid this the telephone company provides velocity equalization which delays the low frequency waves so that all frequencies arrive with a time difference of not more than 10 milliseconds Sound circuits with velocity equalization are not always suitable for television because all frequencies are delayed compared with the picture

Maintenance of program levels It is necessary to maintain program transmission levels within predetermined limits to overcome circuit noise avoid overloading of amplifiers and other circuits and to allow switching of program circuits flexibly and freely without level adjustments This is accomplished by careful preadjustment and continuous monitoring of levels on level indicators, technically known as VU (volume unit) meters. A standard type is used throughout the system to assure uniformity in the observations *See* RADIO BROADCASTING [R.F.C.]

Radio range

A radio facility emitting signals which when received by appropriate companion equipment, provide a direct indication of the bearing of the facility from the vehicle

A N radio range. First installed in the United States in 1927 to form the backbone of the airways system, this equipment was considered obsolescent in 1946 but has continued in use much beyond that date It operated in the 200 to 400 kilocycle range and was received on a high selectivity receiver of conventional design The A N range establishes four radial lines of position which can be identified by a continuous tone signal This continuous signal is actually made up of keyed pulses of equal amplitude representing the Morse code letters A and N The pulses interlock to form a continuous tone on the lines of position, as shown in Fig 1

The principle of this device is illustrated in Fig. 2 This figure represents two loop antennas, which are connected to a radio transmitter by a motor driven switch The transmitter is continuously modulated by a tone to which the ear is sensitive. It is connected to one loop for 3 sec to a second loop for 1 sec, back to the first loop for 1 sec, and back again to the second loop for 3 sec If the receiver is located on a line with the plane of one loop the Morse letter A (dot dash) may be heard

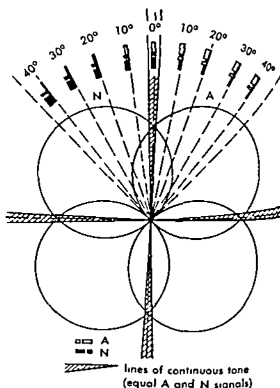


Fig 1 A N radio range signals

mplification must take place between the input of he receiver and its output. This is usually called he gain of the receiver It may amount to 10 000 00 times in voltage or 140 decibels (db) *See* AMPLIFIER

If the detector which abstracts the desired communication from the high frequency amplified signals requires 1 volt to perform its function properly and if the input to the receiver is 1 microvolt a total amplification of 1 000 000 times is required prior to detection If the loudspeaker requires 10 volts another voltage amplification of 10 is necessary between the detector and the loudspeaker

Figure 1 shows the gain between stages of a superheterodyne receiver The voltage gain of any stage or group of stages can be obtained by taking the ratio of the ordinates of the proper curves The gain in *decibels is found* by subtracting the values of the proper curves For example the voltage gain at 1000 kc of the second intermediate-

119–84 = 35

Detection The energy collected by the antenna nd presented to the input of the *receiver is in the* orm of radio-frequency waves which act *as a carrier* for the information to be transmitted The purpose of the detector in a receiver is to remove the desired communication from this *carrier* and to convert it into a form that will actuate the output device such as a loudspeaker See DETECTOR

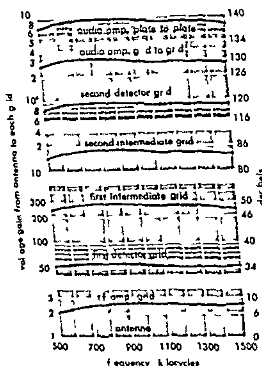


Fig 1 Voltage gain in successive stages of superheterodyne receiver From K Henney ed Radio Engineering Handbook 5th ed McGraw-Hill 1959)

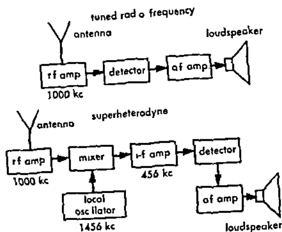


Fig 2 Comparison of TRF and superheterodyne receivers

Types of receivers Two general types of receiver are in use today the tuned radio frequency (TRF) and the superheterodyne *Both of these* can be used for amplitude modulated (AM) signals Frequency modulation (FM) receivers are almost always superheterodyne

TRF receiver In a TRF receiver all amplification up to the detector takes place at the fre

ing elements are ganged together for convenience when tuning For a discussion of tuned amplifiers *see* AMPLIFIER

TRF receivers are especially applicable to the low frequency and very low frequency bands from about 10 000 to 300 000 cps Figure 2 shows a block diagram of a TRF receiver compared to a superheterodyne receiver

Superheterodyne receiver With the increased use of higher frequencies for broadcasting and communications a high degree of selectivity became necessary The TRF receiver was inadequate in this respect This selectivity was obtained by circuit techniques involving frequency changing The superheterodyne circuit changes the frequency by heterodyning or beating two frequencies together to get a third When two signals of small difference in frequency are superimposed on a nonlinear device the output consists of energy at two major combination frequencies which are the sum and difference of the original frequencies The sum o the two beating frequencies is usually eliminate by tuned circuits following the heterodyne process The difference frequency referred to as the intermediate frequency is passed on for further amplification and what is more important through stages of high selectivity because selectivity lower frequencies is more easily obtained than high frequencies

The new i f signal derived from the modulat carrier frequency and an unmodulated local oscillator frequency (see Fig 2) is modulated to t same degree as the original carrier There is

rent in the other antennas are rotated as a function of time Keying of the Morse letters E and T is accomplished by reversing the phases of the current in the outer antennas by 180°

The space pattern of the Consol system is shown in Fig 4 Coverage is provided over an area of about 120° on either side of the antenna array The usual cycle of operation of a Consol station is as follows (1) omnidirectional transmission with identifying station letters for 28 sec, (2) break for 1 5 sec (3) transmission of 30 each of the E and T characters while the lobes rotate for 30 sec, (4) break for 1 5 sec

The Consol signals are received on a conventional receiver operating at low frequencies The pilot counts the number of Es (dots) and Ts (dashes) heard during one cycle and the ratio of these two counts (when referred to a chart) gives bearing to an accuracy of about ½° It is evident from the pattern that Consol has an ambiguity every 40° In order for the Consol information to be useful, it is therefore necessary to have means (such as a direction finder) for determining position to within 20° before this information has significance

Navaglobe. This long range continuous wave low frequency radio range system of the amplitude comparison type provides bearing information for the Navarho polar coordinate system (*see* NAVARHO)

Vhf omnidirectional radio range (VOR). This type of range operates at very high frequencies and provides radial lines of position in any direction

This facility operates in the 112–118 Mc range and employs a directional and a nondirectional antenna The emission from the directional antenna when combined with radiation from the nondirectional antenna, results in a cardioid space pattern This space pattern is rotated at a speed of 30 cycles per second either physically or through the use of a goniometer No other modulation appears on the rotating (variable phase) pattern The rotation is synchronous with a frequency modulation which is imposed on a small amount of radio frequency energy derived from the source that feeds the variable phase antenna and radiated by the nondirectional antenna The output of the nondirectional antenna is therefore a carrier frequency that is amplitude modulated with frequencies that vary cyclically from 9480 to 10,440 cycles at a 30-cycle rate This signal appears everywhere in space (see Fig 5) within the coverage of the station

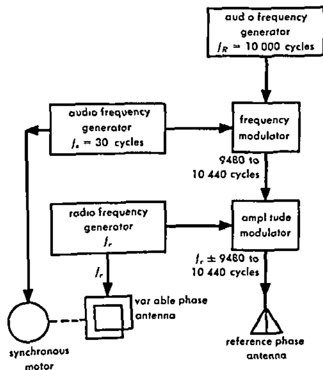


Fig 5 Principle of the vhf omnidirectional radio range

A special receiver is necessary to derive navigational information from the VOR This equipment receives both the variable and reference phase emissions on a common channel After the detector output, a band pass filter operating from 9480 to 10,440 cycles recovers the reference phase signal, while a 30 cycle low pass filter recovers the variable phase signal Phase detectors operating servomechanisms may then be used to indicate bearing. It is also common practice to adjust a phase shifter to a bearing that it is desired to fly The phase detector then indicates when the phase of the received signal is equal to the desired bearing *See* NAVIGATION SYSTFMS, ELECTRONIC [P G S]

Bibliography P C Sandretto *Electronic Navigation Engineering* 1958

Radio receiver

That part of a radio communication system which abstracts the desired information from the radio-frequency (rf) energy collected by the antenna. All radio receivers must perform three basic functions selectivity amplification, and detection For basic discussion of radio principles, *see* RADIO

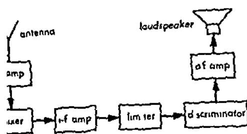
Selectivity. There are many radio signals transmitted at the same time which are available at the antenna Out of these many signals the receiver must select the single one desired This is done by tuning the receiver to the frequency of the desired carrier The tuning circuit contains a combination of inductances L and capacitances C, one or more of which are variable The frequency f selected is determined by the relation

$$f = 1/2\pi\sqrt{LC}$$

where f is in cycles per second (cps), L is in henries and C is in farads

Tuning the receiver therefore consists in changing the inductance or capacitance usually the latter When so tuned the inductance capacitance circuit accepts the desired frequency and rejects other frequencies By using several such circuits in series a high degree of selectivity may be obtained

Amplification. Because the incoming signal may be weak and because a certain minimum energy is required to operate the loudspeaker, the head phones or the television picture tube considerable



3 Block diagram of FM superheterodyne receiver

spectrum conservation and power gain For the ory of single sideband see AMPLITUDE MODULA
...

There are two forms of SSB receivers One mmonly referred to as single sideband (SSB) is the reception of but one sideband of intelli nce on a reduced carrier The other independ t sideband receiver (ISB) is used for the re ption of two channels of intelligence one on an per and one on a lower sideband To effect this pe of reception without undue distortion the car er is exalted or reinforced and in some instances local carrier synchronized by the incoming re uced carrier is employed for demodulation some mes referred to as product detection

To separate the sidebands and carrier carefully esigned filters are employed Their characteristics re very important to reduce crosstalk between he sidebands and the sidebands and carrier chan el

Independent sideband (ISB) enables a great leal of intelligence to be received over a single rf carrier It is possible to receive four telephone con versations in a 12 kc total bandwidth or a single telephone conversation facsimile transmission and many telegraph subcarriers at one and the same time

Diversity reception Signals received at a given point vary in strength (fade) because they are re flected down to the antenna from the ionosphere which is unstable with time Signals at two loca tions spaced a wavelength or more apart do not fade the same amount at the same time They may be strong at one site and weak at the other The correlation decreases with the number of wave lengths ten wavelengths being a practical limit

Diversity reception takes advantage of this phe nomenon Antennas are located at several sites and the detected outputs derived from receivers con nected to these antennas are applied through switching devices to a common output Automatic gating or selection means controlled by the strong est or best signal assure it to be the major or only contributor to the final output Thus the out put will have the best signal to-noise ratio for any combination of operating conditions The final out put of the diversity receiver is derived from the voltage induced on one antenna onl

A high signal to-noise ratio is desirable This may not always be feasible as in the case of re ceivers used in homes for entertainment purposes, ... antenna would be an

...

nications however, a properly desig ...
will reduce the requirement (and expense) of a receiver of low noise factor A good antenna of narrow directivity will extract a large amount of power from the wanted radio wave

The form that antennas take is determined by many factors, according to the particular require ments of the receiving instrument with which they are to be associated the frequency coverage ap plicable the service, the availability of land and so forth The field of radio antenna engineering is very large requiring specialists to resolve each design uniquely *See* ANTENNA (AERIAL)

[W L S]

Bibliography K Henney (ed), *Radio Engineer ing Handbook*, 5th ed , 1959, E A Laport, *Radio Antenna Engineering*, 1952, S A Schelkunoff and H T Friis *Antennas—Theory and Practice*, 1952, S Seely *Radio Electronics*, 1956, K R Sturley, *Radio Receiver Design*, 2d ed , 1953

Radio spectrum allocations

The specification of the frequencies of the radio spectrum which are available for use by the various radio services The radio spectrum is the part of the spectrum of electromagnetic radiation lying be tween the frequency limits of approximately 10^4 and 10^{12} cycles per second For purposes of identi fication the radio spectrum is divided into bands differing from adjacent bands by frequency ratios of 10 These bands are identified by the metric wavelength of the shortest waves in each band (such as the metric band or the decimetric band), or by adjective or numerical frequency designators in accordance with the accompanying table The wave length λ in meters is related to the frequency f in cycles per second by the relationship $c = f\lambda$, where c the velocity of propagation of radio waves in space is approximately 3×10^8 meters per sec ond

Allocation of frequency bands. The table pre sents a listing of the major classes of services which are provided with international allocations within the indicated frequency bands Within the United States there is also a national allocation which con forms in general to the international allocation

The numbers of services and the specific alloca tions and provisions relating to them ...

... so that the frequencies allocated have characteristics satisfying the operational require ments of the service with regard to distance, band width and other parameters. Since many services have varied operational requirements allocations to them will be found in several frequency bands

little distortion in this process which takes place in the mixer

The difference frequency (i f) must be high enough so that little response is obtained to the so called image frequency, which is an incoming (but undesired) radio frequency signal whose difference from the local beating oscillator is the same as the desired signal The image signal differs from the desired signal by twice the i f frequency An example of this is a desired signal of 1000 kc and a local oscillator of 1500 kc beating to obtain an intermediate frequency of 500 kc The image frequency of 2000 kc, if allowed by the selectivity of the TRF amplifiers ahead of the heterodyne, will cause interference by developing the same difference of 500 kc when beating with the 1500-kc local oscillator

The tuning of the local oscillator in home entertainment receivers is ganged to the tuning of the TRF amplifiers and thus does not present any more difficulty in manipulation than does a TRF receiver

Where the ultimate in operational characteristics is desirable it may be required to heterodyne the incoming signal more than once Commercial radiotelegraph traffic handling systems employ at least dual and even triple superheterodyne receivers The first intermediate frequency is made rather high for good image suppression The second intermediate frequency is lowered to a value where other spurious signals are not too obtrusive A third very low intermediate frequency for very high selectivity is then utilized for final detection or demodulation to the original intelligence

The actual change of frequency in a superheterodyne receiver is performed by a frequency converter, often called a mixer, heterodyne modulator, or first detector This device may be a tube transistor or other nonlinear device Two inputs are applied to the mixer the incoming signal and the output of the local oscillator The frequency of this oscillator differs from that of the incoming signal by the intermediate frequency

The mixer may be very simple with both signals applied to the same grid of a tube or it may be more complex It may contain means for generating the local frequency or it may have additional internal elements to which the local oscillator output is applied In any case the mixing process takes place because of the ability which each tube element has to modulate the electron stream from cathode to anode

The greater the number of electrodes in a common electron stream the greater will be the noise developed in the output Therefore multielement mixers are used only in those circuits which have considerable signal amplification ahead of the mixer This reduces the relative noise contribution of the mixer The use of a separate local oscillator allows better frequency stabilization Good frequency stabilization is especially required when the following i f amplifier has a high selectivity (narrow bandwidth)

Local oscillators used in superheterodyne receivers require careful design Assuming that the transmitted signal is kept within narrow limits of frequency tolerance, the local oscillator must keep the resultant intermediate frequency at the center of the pass band of the amplifier This is important to reduce distortion and operator attention.

The narrower the pass band of the amplifier—that is, the greater its selectivity—the more important it is that the local oscillator does not vary in frequency This is difficult in receivers which must be tuned by an operator so that one of several signals may be selected at will In dual detection superheterodynes the second oscillator may be accurately controlled in frequency by a piezoelectric quartz crystal, such as that used to maintain radio transmitters on their assigned frequencies.

Regenerative receiver A very simple and effective form of receiver, often employed in the early days of radio communication, utilizes the phenomenon of regeneration to improve signal strength. In this system, some of the received energy is fed back into the input after amplification If the feedback has the proper phase, the energy fed back adds to the incoming signal and produces a greater output than if no feedback were employed. Although the amplification is high in such a system, the selectivity is not enhanced and for this reason regenerative receivers are seldom employed now

FM receivers. Receivers for frequency modulation systems differ in several respects from those used in amplitude modulation systems

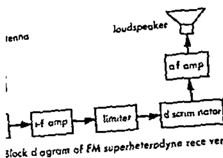
In an AM system the desired communication is impressed on a high frequency carrier by varying the magnitude of the carrier in accordance with the magnitude of the signal to be transmitted The detector in a receiver for this system produces an audible signal corresponding to these amplitude variations

In an FM system the carrier is modulated by the desired message by varying the frequency of the carrier instead of varying its amplitude (*see* FREQUENCY MODULATION) The receiver for such a system must have some means of producing a varying voltage amplitude to correspond to the varying frequency In other words the FM signal must be converted to an AM signal

Because the actual incoming frequency varies the bandwidth to be passed by the receiver circuits must be wide and the greater the frequency variation for a given input voltage variation the greater will be the advantage of the FM system compared to an AM system from the standpoint of eliminating noise

In an FM receiver, a limiter is employed to eliminate all amplitude variations of the carrier and to deliver to the final detector a signal which is free of noise Several types of FM detectors are employed usually called discriminators *See* FREQUENCY MODULATION DETECTOR

Single sideband (SSB) receivers. This type of communication system is advantageous compared to frequency modulation and amplitude modulation



Block diagram of FM superheterodyne receiver

trum conservation and power gain. For the of single-sideband *see* AMPLITUDE MODULA

e are two forms of SSB receivers One nly referred to as single-sideband (SSB) is e reception of but one sideband of intelli on a reduced carrier The other independ leband receiver (ISB) is used for the re- n of two channels of intelligence one on an and one on a lower sideband To effect this of reception without undue distortion the car s exalted or reinforced, and in some instances al carrier synchronized by the incoming re- d carrier is employed for demodulation some- s referred to as product detection

o separate the sidebands and carrier carefully gned filters are employed Their characteristics very important to reduce crosstalk between s debands and the sidebands and carrier chan

ndependent sideband (ISB) enables a great l of intelligence to be received over a single rf rrier It is possible to receive four telephone con ations in a 12 kc total bandwidth, or a single ephone conversation, facsimile transmission, and any telegraph subcarriers at one and the same me

Diversity reception Signals received at a given oint vary in strength (fade) because they are re- ected down to the antenna from the ionosphere which is unstable with time Signals at two loca ions spaced a wavelength or more apart do not fade the same amount at the same time They may be strong at one site and weak at the other The correlation decreases with the number of wave- lengths ten wavelengths being a practical limit

Diversity reception takes advantage of this phe- nomenon Antennas are located at several sites and the detected outputs derived from receivers con nected to these antennas are applied through sw tching devices to a common output Automatic gating or selection means controlled by the strong est or best signal assure it to be the major or only contributor to the final output Thus the out put will have the best signal-to-noi e ratio for any combination of operating conditions The final out put of the d versity receiver is derived from the voltage induced on one antenna only

Receiver antennas The main objective in the des gn of a receiving antenna is extraction of maxi mum power with least noise and interference from unwanted s gnals

A high signal-to-noise ratio is desirable This may not always be feasible as in the case of re- ceivers used in homes for entertainment purposes, where a properly designed antenna would be an inconvenience not required in areas served by high field strengths. In the field of commercial commu nications however a properly designed antenna will reduce the requirement (and expense) of a receiver of low noise factor A good antenna of narrow directivity will extract a large amount of power from the wanted radio wave

The form that antennas take is determined by many factors according to the particular require- ments of the receiving instrument with which they are to be associated the frequency coverage ap- plicable the service, the availability of land, and so forth. The field of radio antenna engineering is very large requiring speciali ts to resolve each design uniquely *See* ANTENNA (AERIAL)

[W.L.S.]

Bibliography K. Henney (ed.) *Radio Engineer ing Handbook* 5th ed 1959, E. A Laport *Radio Antenna Engineering* 1952 S A Schelkunoff and H T Friis, *Antennas—Theory and Practice* 1952 S Seely *Radio Electronics* 1956 K. R Sturley *Radio Receiver Design* 2d ed., 1953.

Radio spectrum allocations

[illegible]

spectrum of electromagnetic radiation lying be- tween the frequency limits of approximately 10^4 and 10^{11} cycles per second For purposes of identi fication, the radio spectrum is divided into bands differing from adjacent bands by frequency ratios of 10 These bands are identified by the metric wavelength of the shortest waves in each band (such as the metric band or the decimetric band) or by adjective or numerical frequency designators in accordance with the accompanying table The wave- length λ in meters is related to the frequency f in cycles per second by the relationship $c = f\lambda$ where c the velocity of propagation of radio waves in space, is approximately 3×10^8 meters per sec ond

Allocation of frequency bands. The table pre- sents a listing of the major classes of services which are provided with international allocations within the indicated frequency bands Within the United States there is also a national allocation which con forms in general to the international allocation.

The numbers of services and the specific alloca tions and provisions relating to them are extremely complex and no complete table of services alloca tions and sharing arrangements can be given here The characteristics of radio waves vary greatly with frequency so that the frequencies allocated have characteristics satisfying the operational require ments of the service with regard to distance, band width, and other parameters Since many services have varied operational requirements, allocations to them will be found in several frequency bands

little distortion in this process which takes place in the mixer

The difference frequency (i f) must be high enough so that little response is obtained to the so called image frequency, which is an incoming (but undesired) radio frequency signal whose difference from the local beating oscillator is the same as the desired signal The image signal differs from the desired signal by twice the i f frequency An example of this is a desired signal of 1000 kc and a local oscillator of 1500 kc beating to obtain an intermediate frequency of 500 kc The image frequency of 2000 kc, if allowed by the selectivity of the TRF amplifiers ahead of the heterodyne, will cause interference by developing the same difference of 500 kc when beating with the 1500 kc local oscillator

The tuning of the local oscillator in home entertainment receivers is ganged to the tuning of the TRF amplifiers and thus does not present any more difficulty in manipulation than does a TRF receiver

Where the ultimate in operational characteristics is desirable, it may be required to heterodyne the incoming signal more than once Commercial, radiotelegraph, traffic handling systems employ at least dual and even triple superheterodyne receivers The first intermediate frequency is made rather high for good image suppression The second intermediate frequency is lowered to a value where other, spurious signals are not too obtrusive A third very low intermediate frequency, for very high selectivity, is then utilized for final detection or demodulation to the original intelligence

The actual change of frequency in a superheterodyne receiver is performed by a frequency converter, often called a mixer, heterodyne modulator, or first detector This device may be a tube transistor or other nonlinear device Two inputs are applied to the mixer, the incoming signal, and the output of the local oscillator The frequency of this oscillator differs from that of the incoming signal by the intermediate frequency

The mixer may be very simple with both signals applied to the same grid of a tube or it may be more complex It may contain means for generating the local frequency, or it may have additional internal elements to which the local oscillator output is applied In any case the mixing process takes place because of the ability which each tube element has to modulate the electron stream from cathode to anode

The greater the number of electrodes in a common electron stream, the greater will be the noise developed in the output Therefore multielement mixers are used only in those circuits which have considerable signal amplification ahead of the mixer This reduces the relative noise contribution of the mixer The use of a separate local oscillator allows better frequency stabilization Good frequency stabilization is especially required when the following i f amplifier has a high selectivity (narrow bandwidth)

Local oscillators used in superheterodyne receivers require careful design Assuming that the transmitted signal is kept within narrow limits of frequency tolerance, the local oscillator must keep the resultant intermediate frequency at the center of the pass band of the amplifier This is important to reduce distortion and operator attention.

The narrower the pass band of the amplifier—that is, the greater its selectivity—the more important it is that the local oscillator does not vary in frequency This is difficult in receivers which must be tuned by an operator so that one of several signals may be selected at will In dual detection superheterodynes the second oscillator may be accurately controlled in frequency by a piezoelectric quartz crystal, such as that used to maintain radio transmitters on their assigned frequencies

Regenerative receiver A very simple and effective form of receiver, often employed in the early days of radio communication, utilizes the phenomenon of regeneration to improve signal strength. In this system, some of the received energy is fed back into the input after amplification If the feedback has the proper phase the energy fed back adds to the incoming signal and produces a greater output than if no feedback were employed Although the amplification is high in such a system, the selectivity is not enhanced and for this reason regenerative receivers are seldom employed now

FM receivers. Receivers for frequency modulation systems differ in several respects from those used in amplitude modulation systems

In an AM system, the desired communication is impressed on a high frequency carrier by varying the magnitude of the carrier in accordance with the magnitude of the signal to be transmitted The detector in a receiver for this system produces an audible signal corresponding to these amplitude variations

In an FM system the carrier is modulated by

frequency In other words the FM signal must be converted to an AM signal

Because the actual incoming frequency varies the bandwidth to be passed by the receiver circuits must be wide and the greater the frequency variation for a given input voltage variation the greater will be the advantage of the FM system compared to an AM system from the standpoint of eliminating noise

In an FM receiver a limiter is employed to eliminate all amplitude variations of the carrier and to deliver to the final detector a signal which is free of noise Several types of FM detectors are employed usually called discriminators *See* FREQUENCY MODULATION DETECTOR

Single-sideband (SSB) *receivers.* This type of communication system is advantageous compared to frequency modulation and amplitude modulation

Inter---ional control The organization for the control radio ion (I) have ratified the Telecommunication Convention (treaty) and its appended regulations relating to radio and cable operations. A Secretariat and ... ency Registration Board

of th ...

groups

Radio Consultative Committee (...) ... technical studies of ways to obtain improvements in systems operation and in spectrum utilization, the International Civil Aviation Organization (ICAO) assists in regard to matters of concern to aviation. The radio regulations divide the world into three regions and contain a detailed table of allocations with certain differences in the three regions. The regulations provide for regional agreements. National allocations can be made within the world wide framework but any departures must be accomplished on a basis of noninterference with the internationally agreed allocations.

United States control Within the United States, the responsibility for frequency allocations is di... on the President who controls fre-

...ntities. Each is empowered to authorize ... any frequency. The Interdepartment Radio Advisory Committee (IRAC), a cooperative group of g...rnment agencies, has been delegated to act for

... rules for the operation of radio stations. This plan is not ideal. The demand for radio assignments has always exceeded the supply. Many of the services do not operate on frequencies which are optimum for the service, the optimum frequencies having been preempted by prior services or techniques not having been developed to make the optimum frequencies available when the service was established or for other causes. A large factor in the problem is the economic cost of making service changes to improve spectrum usage. In addition to the continuing studies being made by the appropriate organizations at national and international levels to improve the allocations, several special studies of the problem have been made by independent groups in government and industry. [E W A]

Bibliography. International Telecommunication Union (ITU), Geneva, *Final Acts of the International Telecommunication and Radio Conferences* ... *munication Convention, Buenos Aires 1952*; Joint Technical Advisory Committee, *Radio Spectrum Conservation*, 1952; U.S. Communications Act of 1934 as amended; U.S. Federal Communications Commission Rules and Regulations Concerning Radio, Part 2; U.S. President's Communications Policy Board, *Telecommunications, A Program for Progress*, 1951.

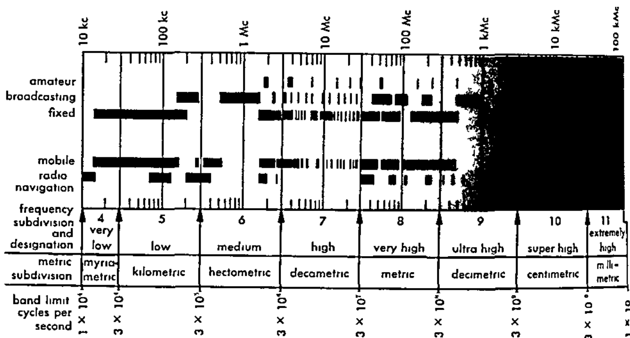
Radio telescope

A *radio antenna* or system of antennas with associated receiving and recording equipment, designed for use in receiving radio-frequency waves from beyond the Earth's atmosphere. Extraterrestrial radio waves were discovered by K. G. Jansky in 1932 with an antenna and a receiver designed to study atmospheric radio static. The first antenna designed specifically for astronomical observations was a 30-ft paraboloidal reflector built in Illinois by Grote Reber in 1937.

Diffraction limitation. The principal function of a radio telescope is the same as that of an optical *telescope*—to collect and concentrate weak elec... from outer space in sufficient quan... dio t... have been ... ability of a telescope to select waves arriving from a given direction while rejecting those from other directions. The ability to separate, or resolve, waves from two bright objects subtending a small angle in the field of view requires a telescope with a size or an over-all extension measured in units of a wavelength equal to the reciprocal of this angle expressed in radians.

The wave diffraction limitation on angular resolving power is a much more serious limitation for radio telescopes than for optical ones because radio wavelengths are 10^{4}–10 times longer. Radio telescopes have diameters approaching 1000 wavelengths, whereas the largest optical telescope has a diameter of 10^{7} wavelengths. Diffraction limits the resolving power only of small optical telescopes. The resolution of large ground-based optical telescopes is seldom finer than about 1 sec of arc because of variations in light ray paths through the Earth's nonuniform atmosphere, an effect which is observed as scintillation of stars (see TWINKLING STARS). This limitation can be overcome only by placing the telescope above the atmosphere, in which case a 15-in. optical telescope would have as fine a resolution as the largest ground-based instrument. Astronomical radio measurements are also ultimately limited by variations in propagation through the Earth's atmosphere and ionosphere.

Choice of location. The radio telescope is relatively poor in rejecting unwanted wide-angle, stray radiation; a rejection factor of 10,000 is not easy to obtain, whereas special optical telescopes are 1000 times better than this. Radio emission from the Earth's atmosphere and ionosphere has a negligible influence at meter wavelengths on the performance of a radio telescope, but man-made signals and natural static arriving at the antenna



International radio frequency allocations

In most of the frequency bands it has been found practical to make block allocations of frequencies to given services Individual frequency assignments of radio stations operating in that service can then be made on a routine basis with little need for coordination with frequency assignments to stations in other services Some of these bands are exclusive to a particular service and some are shared between specified services in varying degrees Similarly, in some services, the stations have individual frequency assignments which are protected from interference by other stations whereas the stations in other services share time on frequencies common to several stations These arrangements are based upon the nature and needs of the services concerned

The safety services are given the greatest measure of protection These include radio navigation marine and aviation Clear channels are generally assigned to the services directly affecting the public, such as common carrier and broadcasting, and services which operate continuously and automatically without the assistance of a professional operator, such as microwave services Services which are adjuncts to various commercial and industrial operations can tolerate varying amounts of interruption or interference and these stations operate generally on shared frequencies

Assignment of specific frequencies. In making specific radio frequency assignments the separation between frequencies which are assigned must be sufficient to provide for the width of transmitted band required for the type of service involved in each case for the separation between transmitted bands necessary to avoid interference and for the lack of perfect stability of the frequency of the radio transmitter. As a result the separation between adjacent assigned radio frequencies varies widely, depending on the type of service and location in the frequency scale

The high frequency band, by reason of systematic reflection of radio waves from the ionospheric layers of the earth's atmosphere, provides international fixed radio communications, international broadcasting and long distance mobile communications for ships and aircraft The characteristics of these frequencies are highly variable Different frequencies within the band have different distance ranges at a particular time, and a particular frequency will have different range capabilities at different times of the day and night, at different seasons of the year and at different epochs of the solar radiation cycle Thus, within this band, a group of spaced frequencies (sometimes called a family of frequencies) is used by stations requiring continuous communications over fixed distances The peculiar characteristics of this range of frequencies results in its great demand by a number of services and the use of large numbers of frequencies by each service Severe congestion results so that a great amount of coordination is required both nationally and internationally Frequency conserving techniques ar [illegible] tions against [illegible] by fixed static [illegible] nication are available limitations on the amounts of power radiated by stations the use of directional transmitting and receiving antennas the use of efficient modulation techniques and bandwidth conserving techniques such as single-sideband operation and time sharing arrangements *See* BANDWIDTH REQUIREMENTS (COMMUNICATIONS)

Control agencies Since the uses and effects of these frequencies are international in scope, an international organization is required for effective management The need for international cooperation is enhanced by the use of frequencies in this and other bands by ships and aircraft which move in international commerce

sion and at centimeter wavelengths by lack of sensitivity

... telescopes Because of the wide

reflector with the diameter ranging ... These are frequently on two-axis equatorial mounts Occasionally two or more up to 32 similar reflectors are interconnected in a system to obtain accurate source size or position measurements by exploiting the resultant complex antenna pattern

... lengths large paraboloidal and

is 250 ft in diameter and is located ... ter England It can operate efficiently at a wavelength of several decimeters A precise radio telescope is shown in Fig 1 It is mounted equatorially on a polar gear 48 ft in diameter The reflector is 85 ft in diameter and has a solid surface accurate to about ⅛ in and therefore can be used at a wavelength of a few centimeters to produce a beam less than 0 1°

A reflector 600 ft in diameter on an altitude-azimuth mount, shown in Fig 2 is being built in West Virginia for the U S Naval Research Laboratory to be completed in 1962

By electronically comparing the output signals from two linear array antennas perpendicular to each other a single effective antenna beam of circular cross section is produced This is called a Mills cross and is used at meter wavelengths in mapping the radio intensity of the celestial sphere and in recording the position and flux of radio sources.

The angular resolving power of a large antenna (in one example it is 64 acres) can be obtained with much smaller antennas by synthesizing the large effective area mathematically from the daily output records obtained from two small antennas connected as interferometers as a belt of the sky passes through the interference pattern Each day the separation and the azimuth of the base line between the antennas is varied until all independent arrangements have been covered A computer then calculates with high angular resolution the entire radio-intensity distribution over the fraction of the sky surveyed See RADIO ASTRONOMY

[F T H]

Bibliography R H Brown and R Q Twiss A new type of interferometer for use in radio astronomy Phil Mag 45 663–682 1954 J D Kraus Radio telescopes Sci American 192 36–43 1955 A C B Lovell Radio astronomy and the Jodrell Bank telescope Proc IFE pt B vol 103 711 721 1956 L Marton (ed) Advances in Electronics and Electron Physics, vol 7 1955 G S Mumford, III Radio observatories of the world Sky and Telescope vol 18 1959

Radio transmitter

A generator of high frequency electric current whose characteristics of amplitude frequency, or phase angle may be altered or modulated in accordance with the intelligence to be transmitted A radio transmitter consists of several distinct major components to accomplish the objectives of a particular design for a particular service requirement.

The power a transmitter delivers to the antenna may vary from a fraction of a watt to 1 000 000 watts. Lower powers are used mainly for portable or mobile services while higher powers are required for broadcasting over large areas and in point to-point communications

Transmitters may be classified by the type of modulation used Amplitude modulation (AM) transmitters are employed for broadcast purposes at medium frequencies Frequency modulation (FM) and phase modulation require much larger bandwidths in broadcast service and are used mainly at very high frequencies for broadcast purposes. Frequency and phase modulation provide greater signal to-noise ratio than amplitude modulation for the same antenna input power There is also an advantage in operating at very high frequencies where noise is considerably less than at the lower or medium frequency band Single-sideband (SSB) or independent sideband (ISB) trans...

... ble in comparison to the total power of the transmitter The modulation of the transmitter in this mode is both amplitude and angular The principal application of single-sideband transmission is for point to-point long distance telephone and telegraph circuits A particular type of SSB transmission called compatible single sideband may be utilized for broadcast program transmission since it can be received by the usual AM broadcast receiver See AMPLITUDE MODULATION FREQUENCY MODULATION PHASE MODULATION

Amplitude modulation transmitters AM transmitters have two principal design types either low level modulated or high level modulated The low level modulated transmitter is modulated at its low power stages requiring little modulation power The high level modulated transmitter which usually accomplishes modulation at the anode of the output power amplifier requires modulation power to be equal to approximately 50% of the power of the carrier

In order to amplify faithfully and reproduce the modulation at low levels the power amplifiers must be linear These are usually Class B linear amplifiers which are much less efficient than the Class C amplifiers employed in high level modulated carrier amplifiers Because it operates over a large band of frequencies the high power modulator must have linear power amplification to achieve low distortion See AMPLIFIER

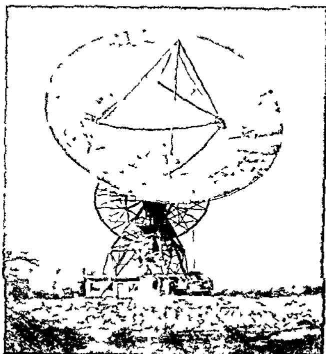


Fig 1 Precision radio telescope 85 ft in diameter on an equatorial mount This reflector can selectively receive centimeter radio waves from an angular area of only 3 5% of the face of the Moon (U S Office of Naval Research and University of Michigan)

from nearby or from beyond the horizon seriously interfere with accurate measurements and weak signal detection Consequently, radio telescopes are most favorably located far from populated areas in valleys between mountain ranges The mountains shield the antenna from ground wave radio interference, bad weather, and damaging winds In contrast optical observatories are placed on mountain tops to avoid much of the turbulent and impure lower atmosphere

At short radio wavelengths it is convenient to mount the antenna in a manner similar to that of an optical telescope, thereby making it easy to track celestial objects across the sky with a simple mechanical drive Although the required precision of angular pointing of a radio telescope is much less than that of an optical telescope the problem is still difficult for the larger antennas because of their size

Special radio telescopes have been built to operate over a wide band of wavelengths simultaneously without loss of performance, other antennas are designed for the measurement of the state and degree of polarization of the incident solar, planetary, or cosmic radio waves

The detector of a radio telescope is normally a radio receiver (although at millimeter wavelengths, heat detectors have been employed), superheterodyne, tuned radio-frequency, or traveling wave tube receivers, with or without solid state low noise amplifiers, have been used The output signal from the receiver is recorded on continuously moving paper tape, magnetic tape, or photographic film The observing and recording techniques in radio and photoelectric astronomy have much in common

Discrimination between signal sources. The maximum number of cosmic radio sources that can be reliably measured both in position and in intensity with a given radio telescope is determined either by confusion between sources due to inadequate resolving power or by lack of sensitivity due to inadequate signal to noise ratio The antenna directive gain is the ratio of 4π steradians, the solid angle of the celestial sphere, to the solid angle of the antenna beam measured in steradians at the half intensity level. It has been found by experience that the maximum number of discrete radio sources that can be measured reliably is less than one tenth or one twentieth of the antenna gain For example, at a wavelength of 1 m, an antenna diameter of 70 m can produce a 1° beam of circular cross section This telescope with a sensitive receiver is not limited by energy sensitivity but by confusion, and therefore can reliably measure less than 4000 individual radio sources, or an average of 1 source per 10 square degrees of sky

On the other hand, at a wavelength of 10 cm a 1° beam can be obtained with a 7 m diameter antenna But this telescope is limited by lack of sensitivity because of the small collecting area and the fact that most radio sources emit less at shorter wavelengths, with a sensitive receiver, it is impossible to detect more than about 600 radio sources. For comparison, about 10 000 stars per square degree can be detected optically away from the Milky Way In general at meter wavelengths, the large radio telescopes are limited by source confu



Fig 2 Artist's sketch of aluminum mesh reflector and steel mount of Sugar Grove telescope Supporting arcs enable telescope to be tilted in elevation Entire structure will also rotate on circular track on ground (U S Naval Research Laboratory)

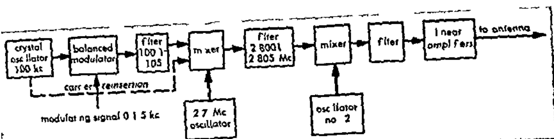


Fig 3 Schematic diagram of SSB transmitter (From F E Terman Electronic and Radio Engineering 4th ed., McGraw-Hill 1955)

amplitude of the transmitted radio frequency is constant, therefore the entire FM system is arranged to be insensitive to amplitude disturbances

FM transmitters are basically similar to AM transmitters except that the AM modulation amplifying system is dispensed with and the exciter must be a variable frequency source One method of modulating the frequency at the exciter utilizes a reactance in the frequency-determining section of the oscillator in the exciter The reactance value is changed electronically or electrically in accordance with the low frequency modulating signal. The remainder of the transmitter as in the AM case, will be made up of frequency doubling and tripling stages of power amplification and a power output stage

Carrier stabilization For FM transmitters this is a more difficult problem than for AM transmitters Many schemes have been used including those used for carrier automatic frequency control in receivers See AUTOMATIC FREQUENCY CONTROL (AFC) Another means of frequency stabilization employs heterodyning a high frequency crystal generated wave with a lower frequency well stabilized tank-circuit oscillator modulated by push pull reactance tubes *See* FREQUENCY MODULATORS.

Protection Protection of FM transmitters is similar to that in AM transmitters except that there are fewer components involved and the system is therefore not as extensive

Monitoring FM transmitter monitoring differs

ulation A frequency discriminator tuned to the relevant frequency must be used to reconstitute the modulation at the transmitter monitoring position

Single sideband independent sideband SSB/ISB transmission promises great growth of communications because of several peculiar system characteristics The ability to transmit intelligence occupying only half the bandwidth as compared to AM and FM modes and its adaptability for modulation by many low frequency sub-carriers are principal advantages

The decreased band occupancy and reduced carrier effectively concentrate all the intelligence in half the bandwidth greatly increasing the effective power in voice or broadcast Decreased bandwidth ... ion doubles the number of services which can use any portion of the frequency spectrum.

Without SSB/ISB means for transmission radiotelegraph and radiotelephone circuits are difficult to operate on closely spaced frequency assignments Maintaining precision and stability at high radio frequencies is much more exacting than at the sub-carrier modulation frequencies in the low audio spectrum

Typical equipment arrangement for SSB transmission comprises an SSB generator (exciter) which makes the SSB or ISB signal (using a system of balanced modulators usually) and filters to separate out the carrier for reinsertion in the desired amount select the desired sideband and reject the undesired sideband The output, at low levels of several watts is then amplified without frequency changing in linear Class B amplifiers to the final power output for coupling to the antenna This method (Fig 3) is commonly referred to as linear SSB transmission A principal design objective is the maximizing of power output for minimum spu

citer or adaptor which splits the SSB signal into its two components of amplitude and phase modulation the normal transmitter stages (which may employ

power

nent)

nent for insertion into the modulator as an ordinary AM signal These two components are then recombined after phase and amplitude equalization at the highest power level This method is called envelope elimination and restoration single-sideband The major design objective here is the reduction of spurious radiation which is independent of the power output

A so-called compatible single-sideband system, utilizing the techniques of the last paragraph modifies the amplitude envelope by using product detection for the AM component and full instead of reduced carrier Compatibility is effected because the amplitude envelope of this SSB wave is similar to that of a normal AM wave and can be received by the usual AM receiver [W LS]

Bibliography L. R Kahn Single sideband transmission by envelope elimination and restoration, *Proc IRE* 40 803-806, 1952 E A Laport

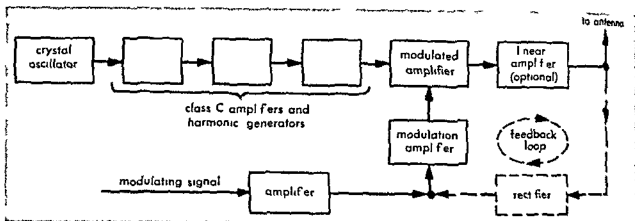


Fig 1 Schematic diagram of AM radiotelephone transmitter (From F E Terman Electronic and Radio Engineering 4th ed McGraw Hill 1955)

Typical equipment A low power rf oscillator whose frequency is very accurately controlled (since it determines the final *carrier frequency*) is the exciter for the transmitter *The exciter* is fol lowed by several stages of *power amplification* which are required to drive the final *power output* stage In low level modulated transmitters all the amplifiers following the *modulated* stage are tuned to the same *frequency* Those ahead of the modulated stage may be used to *double or triple* the frequency of the exciter In a high *level modu* lated transmitter the power amplifier stages are seldom tuned to the same frequency except for the input and output tank circuits of the *power ampli* fier which feeds the antenna

The modulator of a high level *transmitter* de rives its input from a microphone or other source of audio signal and amplifies the signal with low distortion of the order of 1% to a *level which* is usually half that of the carrier power

Most high power modulators utilize *push pull* in either Class A or linear Class B to reduce dis tortion by balancing even harmonics and to balance out some hum and noise *components* *Negative* feedback is also used for this purpose

Antennas Antennas used with transmitters transform the power generated to an electromag netic field They are designed to have a high ratio of radiation to total resistance which determines the efficiency Other factors in the design are a suf ficient bandwidth to accommodate the frequency band transmitted directivity, and the restriction of the solid angle of radiation With certain types of antennas it is possible to suppress radiation of harmonic frequencies which may cause interfer ence for other services It is more usual however to utilize harmonic traps to couple the antenna to the transmitter

Protection This important design objective in transmitters assures continuity of operation and protection of personnel The power supply and *high voltage* system design are probably the major problems for any new transmitter type. Safety in terlocks circuit breakers and warning lights are employed extensively Automatic discharge of ca pacitors for the protection of personnel is impor *tant* For the protection of equipment the tem perature of the cooling water or air draft must be maintained low enough for efficient operation or the equipment is shut down automatically

Monitoring The operation of a transmitter must be *monitored* at all times to keep the operating personnel informed of the condition of transmis sion Meters in important parts of the circuit, os cilloscopes and aural monitoring are all used

Frequency modulation transmitters The trans mitting *frequency* in an FM transmitter is varied above and below the median by an amount accord *ing* to the amplitude of the modulating signal and at a rate determined by the modulating signal The

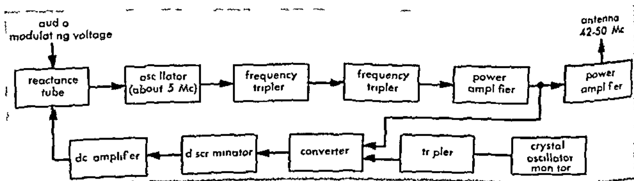


Fig 2 FM reactance tube transmitter (From F E Terman Radio Eng neers Handbook McGraw H ll 1943)

fraction brought down is less than for land bursts, possibly as little as 20%

This early fallout is called local *fallout* because it falls in the vicinity of the bomb *explosion* *Local* fallout may extend downwind 200–300 *miles*, depending on the strength *of the* wind and the bomb yield For example, in March 1954, a megaton yield bomb fired on the Bikini coral *atoll in the* Pacific gave a local fallout over *7000 mi²* as shown in Fig 1 The fine dust made *from the totally and partially vaporized coral* fell back in the downwind direction in the first 36 hours after the explosion Because the local fallout descends so quickly and in *such a concentrated way* it is much more dangerous in the affected areas than are the slower falling types of world wide fallout which constitute the remainder of the fallout

The mixture of fission products formed in a nuclear detonation decays in such a way that the radiation intensity decreases tenfold for every sevenfold increase in age Thus fallout occurring at 1 hour after the explosion will at the end of 7 hours be reduced to 10% of the intensity it had when it first fell and at the end of 49 hours or 2 days and nights its intensity will be 1%

The local fallout being so newly formed and

is in large part the p...

hiding from local fallout Some warning will always be possible so shelter can be sought with good hope of success Fallout shelters can be improvised in many existing buildings the basement of the ordinary home being a good example

Stratospheric and tropospheric fallout The

is truly world wide it circulates widely before it finally descends Current estimates are that it spends an average of about five years in the air if injected near the equator or as little as six months if injected in the fall of the year near the poles although the residence time for the debris from any given injection varies with latitude altitude and time of year Thus its radiations are much reduced in intensity when deposition finally does occur

tons (megatons) of ordinary chemical explosive while atomic bombs equate to thousands of tons (kilotons) Stratospheric fallout results from hydrogen bombs and tropospheric fallout from atomic

he troposphere

No fine particulate matter can remain in the troposphere for more than 1 month or so because rain and snow continually wash it clean The washing mechanism seems to be at least as effective on the smallest particles as it is on the larger ones Apparently the smaller particles are rapidly kicked around in a zigzag path by collisions with molecules of the air, and as a result they have a good chance of hitting the droplets of water which make up clouds and of sticking to them Thus the washing process really is more of a cloud air cleaning mechanism In a matter of 1 month or so essentially all tropospheric air spends a few hours in a cloud, so that the whole troposphere is washed clean in 1 month or less In addition, the fine particles can stick to leaves on trees and on grass and ... Thus tropospheric fallout is carried

only such ...

the fine particles, a minor effect relative to the rain or snow deposition

Because tropospheric fallout lasts only 1 month or so it is restricted latitudinally to the region of the test, a band about 10° wide around the earth This band is more or less uniformly covered except when there are local variations in rainfall

The stratospheric fallout occurs most probably because of gradual mixing of stratospheric air—something like 20% per year for equatorial shots—with the troposphere Here the weather phenomena take control and the fine fallout particles are brought down by rain and snow and surface contact just as for tropospheric fallout

Fallout hazards The principal hazards from fallout are the γ radiation from external fallout, particularly in local fallout and the radiations from internally assimilated fallout principally of the world wide varieties At the present time the latter has been discussed much more widely than the former This is because local fallout is carefully controlled in the testing of nuclear weapons whereas world wide fallout is less readily controlled In nuclear warfare however the external radiation from local fallout probably would be the principal hazard

The γ radiation effects are twofold *the effects* on an individual s health (somatic effect) and the genetic effects on future generations The health effects range *from disabilities* noticeable within a few hours of intense exposure which may lead to death in a few days to the induction of leukemia and possibly cancer of the bone which may appear years later These serious health effects are first obvious at doses of about 100 roentgen units (r) of whole body *radiation* Death occurs for whole body radiation in half the cases of exposure at about 450 r *Natural* radiation from the ground human bodies, and cosmic radiation amounts to a total of 0 1–0 2 r/year depending on locality and altitude

The genetic effects which will develop only in later generations are not known for humans Judging *from* experiments on animals and plants how

Radio Antenna Engineering, 1952, S A Schelkunoff and H T Friis, *Antennas—Theory and Practice*, 1952, S Seely, *Radio Electronics*, 1956

Radioactive fallout

The radioactive debris of nuclear explosions, consisting of the materials formed by the splitting of uranium or plutonium atoms (*see* NUCLEAR EXPLOSION) This material is incorporated into and deposited on the surfaces of the dust formed by the explosion of a nuclear bomb or other nuclear weapon The dust is formed by the vaporization and subsequent recondensation of the solids from bomb parts dust already in the air, and dirt from the ground if the explosion occurs close enough to the surface for the incandescent fireball to touch the ground

Fallout radiation The radioactivity of the fission products is due to the natural instability of the atoms formed by the fission act This instability is removed by the emission of radiations The radiation is either the simultaneous emission of an electron and an antineutrino—a process often called β particle emission or β radiation—or the emission of a γ ray, which is a type of electromagnetic radiation The first process—electron radiation with antineutrinos, the latter being themselves essentially undetectable—constitutes on the average about two thirds of the radiations emitted by fallout, γ rays constitute the remainder *See* FISSION, NUCLEAR, RADIOACTIVITY

An important difference between the two kinds of radiation is that the electron (β) radiation has much lower penetrating power, although they both carry about the same total energy on the average The electron radiation is 50% absorbed in about 1/10 of 1 in of wood flesh, or water, while the γ radiation requires about 2½ in of the same materials for 50% absorption In denser matter the penetrations are less, but the order and relative ranges of penetration are approximately the same as in light material *See* BETA RAYS, GAMMA RAYS

The effects of fallout radiation thus are due to these very different kinds of radiation The short range more absorbable electron radiation is able to cause effects only when the fallout is either taken internally or is in contact with the skin because ordinary clothing is adequate protection against it On the other hand the γ radiation is so penetrating that 1–2 yd of earth or concrete is required for full protection under heavy fallout conditions Gamma radiation requires as many as ten factors of two reduction in intensity for reasonable safety

As stated before, the radioactive fission products produced in the explosions become part of and attach themselves to the solid particles formed by the cooling gases of the fireball part of which are volatilized solids from the bomb, air dust, and (in the case of surface bursts) soil particles After the detonation, these particles return to earth carrying the fission products This is radioactive fallout

Local fallout. The size of the particles is of great importance in determining the rate of return. In the case of surface bursts over land not only is soil vaporized and then recondensed together with the vapor from the bomb itself but great tonnages of soil are also taken up into the fireball and melted and partially vaporized These effects result in larger particles than are formed in the case of air bursts or in the case of surface bursts over sea water Consequently, the fallout falls more rapidly from surface bursts and to a greater degree.

In the case of ground bursts, as much as 80% is brought down in the first few hours in this way in an oval shaped pattern stretched in the downwind direction with intensity contours approximating those shown in Fig 1 Similar early fallout occurs for surface bursts over water, although the total

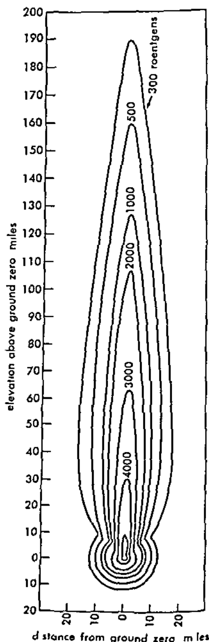


Fig 1 Idealized total (accumulated) dose contours from fallout in first 36 hours after the high yield explosion at Bikini Atoll on March 1, 1954

economic importance are the oxide uraninite and its variety pitchblende the vanadates carnotite and tyuyamunite the silicate coffinite, the phosphates autunite and torbernite and the complex oxides brannerite and davidite The chief thorium minerals of economic interest are the silicates thorite and thorogummite and the oxide thorianite *See* RADIUM THORIUM URANIUM

The minerals that contain uranium or thorium in small amounts as vicarious constituents are relatively numerous Some of them are of present or potential economic interest as sources of uranium or thorium particularly when these elements can be obtained as by products of the recovery of associated elements or minerals The chief source of thorium in the past has been the rare-earth phosphate mineral monazite, which commonly contains ... the range

and is present sometimes in significant amounts in many other minerals including allanite zircon and the apatite of some phosphate rock deposits *Thorium occurs as an accessory* constituent chiefly in minerals containing zirconium cerium calcium *or uranium*

The radioactive decay *of the uranium* and thorium *present* in minerals is accompanied by the emission of α particles electrons and γ radiation (x rays) and a variety of methods of study of such minerals and of prospecting for them *are based on*

or entirety by the internal absorption of *radiation* chiefly α particles released within the crystal In the final stages of radiation damage the crystal may become transformed into an amorphous glassy

chemical alteration Lead accumulates in all radioactive *minerals since it is the end product of the radioactive* decay of the uranium and thorium, and measurement of the ratio of the lead to uranium and thorium or of lead *isotope* ratios has been widely used as a method of measuring geologic time

From the crystallochemical viewpoint *minerals* that contain uranium fall *into two* broad categories those that *contain uranium* in its ...

... as primary or hypogene deposits They include uraninite and coffinite among the *uranium minerals* proper together with virtually all the minerals that contain uranium as a vicarious constituent Most of the known uranium minerals contain the hexavalent uranyl ion instead of quadrivalent uranium These uranyl compounds are characterized by a relatively bright lemon yellow

to green or orange color, and commonly fluoresce a bright lemon yellow color in ultraviolet radiation, the colors may be modified by the nature of other cations present in the mineral in addition to uranium Virtually all of the uranyl minerals are secondary in origin and form from solution at relatively low temperatures and pressures *See* ACTINIDE ELEMENTS, LEAD ISOTOPES GEOCHEMISTRY OF, RADIOACTIVITY, ROCK (AGE DETERMINATION), *see also* ALLANITE, CARNOTITE, MONAZITE, NUCLEAR FUELS, THORIANITE, THORITE, URANINITE, ZIRCON [C.F.R.]

Bibliography H Faul (ed), *Nuclear Geology*, 1954, C. ...
nium and ...
1958, E ...
Radioactive Raw Materials, 1958, K Rankama, *Isotope Geology*, 1954

Radioactive species produced by cosmic rays

A group of naturally radioactive isotopes generally produced by the bombardment of air by *cosmic* radiation These nuclides (atomic nuclei) occur toward the light end of the periodic table, and may have *short half lives* They are observed on earth only as a result of their continuous creation in the atmosphere

Their importance arises chiefly from *their* application as *tracers in* geochemical processes and in age determination By far the most useful is carbon 14 whose combination of half life (about 5600 years) and chemistry makes it ideally suited for studies of the human past (*see* RADIOCARBON DATING) Other important *members* of the group are *hydrogen-3* (*tritium*), beryllium 7 and beryllium 10 *New ones* are still being discovered

Source The cosmic radiation as it comes in from outer space (primary *cosmic* radiation) consists *chiefly* of protons and alpha particles, along with a small number of nuclei of heavier atoms (*see* COSMIC RAYS) Typically the energy of each particle is more than 1 000 000 000 electron volts. This is a very large quantity compared to the few million electron volts with which nuclei are bound together Thus a very violent and extensive reaction can take place when a cosmic ray particle strikes a nucleus The products of such a reaction may themselves possess enough *energy* to cause nuclear reactions, and these in turn may yield still other reactive particles (secondary cosmic radiations) Eventually, however, the energy is distributed among enough particles so that no further nuclear

since this particle can penetrate the nucleus without *repulsion* Its high reactivity is actually enhanced at the lowest *energies* (*see* NEUTRON) Every neutron will cause a further transmutation In the earth s atmosphere one reaction will predominate

ever it is expected that about 10% of the present genetic mutations are due to the natural radiation The radiation genetic effects may be proportional *to total* dose and cumulative that is it is not certain that recovery from radiation induced mutations occurs in general The present pool of human reproductive germ plasm has an enormous accumulation of mutations presumably largely of chemical origin Judging from the estimate of 10% of the natural rate of generation of new mutations for *0 1 to 0 2* r/year however one can estimate the genetic effects of world wide fallout from nuclear weapons testing or from particular types of nuclear war campaigns The irrevocability of such effects has caused much concern although the present effects from nuclear testing are quite small

Strontium 90 Probably the most serious world wide hazard from fallout is strontium 90 one of the fission products produced in highest *yield It* has a relatively long average life of 40 years and a chemical similarity to the element calcium which causes it to fix itself in bone in a semipermanent fashion (*see* STRONTIUM) The bone is exposed to electron radiation from strontium 90 *and its short* lived radioactive daughter product yttrium 90 however neither this isotope *nor its daughter* emits γ radiation and therefore strontium 90 is not a genetic hazard because bone structures *are* not near the reproductive organs The irradiation of the bone structure can cause cancer *of the bone* and leukemia these deleterious health effects constitute the hazard of strontium 90 The fallout is assimilated in milk and dairy *products and in* vegetables—the regular sources of dietary calcium—and the strontium 90 is taken in *from these sources* In 1959 in the populous part of the northern hemisphere new human bone—as in young *children—re*ceived 0 006 to 0 010 r/year from the strontium 90 of past nuclear tests This is very small *relative to* the natural background but in time of nuclear war it could become much larger over a major part of the world (see Fig 2)

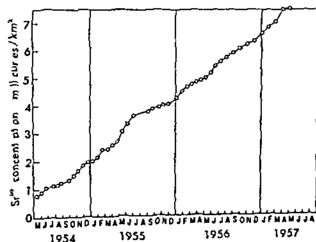


Fig 2 Cumulative deposition of strontium 90 at Milford Haven Wales (After N G Stewart et al in E A Martell *Atmospheric effects of strontium 90 fallout Science* 129 (3357) 1197–1206 1959)

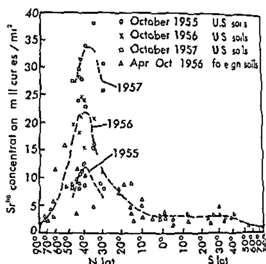


Fig 3 Surface distribution of strontium 90 (E A Martell *Atmospheric effects of strontium 90 fallout Science 129 (3357) 1197–1206 1959*)

The fallout in the Northern Hemisphere has been heavier than in the Southern thus the Southern effects to be expected are proportionately smaller (see Fig 3) Tests on the bones of children from the Southern Hemisphere show only 25% of the amount of strontium 90 found in the bones of children from the United States and northern Europe.

Cesium 137 Another fission product cesium 137 which is a high yield long lived (40-year average life) γ emitter is seriously watched in the world wide fallout Being like sodium chemically it is not permanently fixed in the body but stays only 6 months or so before passing out again Its effects are the same as for external γ radiation The 1959 level was about 0 0015 r/year from the cesium 137 as a result of past nuclear weapons tests The external γ radiation intensity in the Northern Hemisphere is about 0 003 r/year *See* RADIATION INJURY (BIOLOGY) [W F L]

Bibliography H W Feely The Sr^{90} content of the stratosphere *Science* 131(3401) 645–649 1960 Hearings of the Joint Committee on Atomic Energy May 27 June 7 1957 *Radioactive Fallout* pp 554–568 3 vols 1957 J L Kulp A R Schulert an
III Science
Langham Tl
fallout from
ics 1(2) 105–124 1958 W F Libby Worldwide radioactive fallout *Proc Nat Acad Sci* 42 365–390 1956

Radioactive minerals

The so called radioactive minerals loosely comprise species that contain uranium or thorium as an essential part of their chemical composition together with minerals in which these elements are sometimes present in solid solution usually in small and variable amounts About 150 minerals fall into the first category including many that are rare or that are imperfectly known The principal uranium minerals from the *point of view of their*

limit of vertical mixing in the atmosphere) is much lower in altitude at high latitudes and thus the fraction of Be-7 atoms produced n the troposphere is much less than it is near the equator Unlike carbon 14 and tritium Be-7 is an isotope of a rare element. This facilitates its concentration and measurement

Beryllium 10 because of its almost unique half life, midway between carbon 14 and the long lived ...topes holds promise as a useful species for age ...surement in the million year time span It has ...n measured in a number of samples of deep sea ...ments However its actual employment awaits ... answer to some questions on the geochemistry ... this unusual element The existence of Be-7 is ...oving very helpful here

...Some penetration of cosmic ray particles into the ...rth has been observed especially at high eleva...ons It has been suggested that chlorine 36 pro...uced in rocks may under special conditions be ...seful for age measurements

A fruitful field of research now opening up is the ...dy of radioactivity produced in meteorites by

been observed Combining measurements of radio-

come from outer space {J.R.AR}

Bibliography J R Arnold Beryllium 10 produced by cosmic rays *Science* 124 584–585 1956 J R. Arnold and H Al Salih Beryllium 7 produced by cosmic rays *Science* 121 451–453 1955 P S Goel, Radioactive sulphur produced by cosmic rays in rain water *Nature* 178 1458 1956 W F Libby *Radiocarbon Dating* 2d ed 1955

Radioactive waste disposal

The handling and disposal of radioactive wastes is ...roblem present in some degree in all nuclear ...rgy operations Wastes in liquid solid or gase... form are produced in the mining of ore pro...ction of reactor fuel materials reactor opera...n processing of irradiated reactor fuels and a ...eat variety of related operations Wastes also re...lt from the use of radioactive materials for exam...e in research laboratories industrial operations ...d medical research and treatment The prob...ms of waste disposal undoubtedly will increase ... the future as the nuclear energy program is fur...her extended and diversified and as a large and ...despread nuclear power industry is developed

In the handling and disposal of radioactive ...astes the principal problem is the prevention of ...radiation damage to man and his environment by controlling the dispersion of radioactive materials Damage to man may result from irradiation by external sources or by the intake (by ingestion by inhalation or through the skin) of radioactive materials their passage through the respiratory and gastrointestinal tract and their partial incorporation into the body Radioactive waste contaminants in air water food and other elements of the human environ

permissible

...active materials require ultimate disposal in isolated and permanent containment media where they never again can find their way into man's environment The more dangerous radioactive materials are those that may be readily incorporated into the body and that have relatively long half lives ranging from a few years to several thousand years Both the short lived radionuclides and those

radionuclides with extremely long half lives have such low specific activity that is, so few microcuries per gram that the probability of dangerous quantities entering the body is very low Some of the more dangerous radioisotopes include fission products such as strontium 90 and cesium 137 transuranic elements such as plutonium 239 and americium 241 and naturally occurring radionuclides such as radium 226 and thorium 230 See RADIATION INJURY (BIOLOGY)

The highly radioactive liquid wastes associated with chemical reprocessing of reactor fuels constitute the major waste-disposal problem However the liquid solid and gaseous wastes of low or intermediate levels of activity from hospitals, industrial laboratories research reactors and so on must be controlled also and their dispersion limited as necessary to prevent health hazards

The two basic methods for disposal of radioactive wastes are (1) dilution and dispersion and (2) concentration and permanent containment. Where only small amounts of radioactive materials are involved and the local situation is favorable wastes may be diluted and dispersed in water or air without danger In cases where dilution is not feasible wastes must be concentrated and stored in a safe manner and much research and development work has been devoted to methods for concentrating and storing gaseous liquid and solid radioactive wastes

Liquid wastes *High level* liquid wastes which result from experimental or operational processing of irradiated reactor fuels are relatively small in volume but high in specific activity The radioactivity of fuel processing solutions may range from several hundred to thousands of curies per gallon depending upon the processes employed These wastes which may be highly acid or alkaline present extremely difficult problems in shielding, handling and ultimate disposal

High level liquid wastes have been stored in underground tanks of steel and concrete with special

Nitrogen 14 + neutron → Carbon 14 + proton
or $N^{14}(n\ p)C^{14}$

Since about half of nuclear matter consists of neu trons, they are very common products of nuclear reactions at all energies, and the products of the highly specific, low energy reactions they undergo are also very common

The 'debris" remaining after high energy reac tions or series of reactions includes all nuclides whose mass is less than the sum of the masses of the bombarding particle and target nucleus Two large classes of these are virtually undetectable however, under natural conditions First, the stable nuclides produced which must be detected by their properties in bulk are with one or two interesting exceptions masked by the vastly larger amounts of these nuclides already present in the environment Second, the radioactive nuclides of very short half life decay before they can be collected Thus, in general attention is directed to the possible radio active species of moderate or long life

The table lists those nuclides so far discovered which are produced in the earth's atmosphere Be cause nitrogen and oxygen are so much more abundant than argon the rate of production of species made from them is much higher

The primary cosmic rays penetrate only a frac tion of the earth's atmosphere before they undergo a reaction The products may themselves move some distance before they are taken up in turn The most important zone of reaction is at a height of about 10 miles above sea level (although some transmutations occur in every part of the air) Thus a majority of the atoms are produced in the strato sphere and only very few at sea level a fact of im portance in their later distribution For example in the case of carbon 14 only a negligible fraction of the carbon 14 atoms in a sample of living mate rial were made in the sample itself Virtually all are brought in by metabolic processes

Only a little is known directly about the chemi cal behavior of the newly formed radioactive atoms in the atmosphere but much can be inferred The carbon 14 atoms seem to form carbon dioxide rather quickly while the tritium atoms are fixed in water and in molecular hydrogen Only a small fraction of the tritium atoms are in the latter form but since there is only a very small amount of hydrogen in the air it is enough to produce a relatively high specific activity in molecular hydro

Radioactive nuclides produced by cosmic rays in air

	Nuclide	Half life	Radiation, mev
From nitrogen with neutrons	C^{14}	5570y	β 0 16
From nitrogen and oxygen with high energy particles	H^3	12 3y	β 0 017
	Be^7	53d	γ 0 48
	Be^{10}	$2\ 5 \times 10^6$y	β 0 56
From argon	P^{32}	14d	β 1 7
	P^{33}	25d	β 0 25
	S^{35}	87d	β 0 17
	Ar^{37}	35d	K-capture

gen In the case of beryllium and most of the other elements, there are no gaseous compounds As a result, the elements (probably usually as oxides) adhere to the dust of the upper atmosphere They later serve as tracers for this dust

Applications. In principle any natural radioac tive nuclide whose production and distribution are well enough understood can be used for mea uring times comparable with its half life On the assump tion that it has been produced at a steady rate throughout the recent past, and that it is rapidly mixed with the atmosphere and with living matter a highly successful method of age determination has been developed using carbon 14 This nuclide has also been applied to studies of the rate of mix ing of the oceans and the rate of transfer of carbon dioxide between air and ocean The decrease in the C^{14} concentration of atmospheric and living carbon in the last century because of dilution with inactive carbon dioxide from the burning of fossil fuels and the production of carbon 14 by thermonuclear ex plosions have also been useful in this way Precise measurements are necessary for this purpose.

Tritium is a constituent of a very important sub stance water, and helps to trace the path of circu lation of surface waters Because its half life is so much shorter than that of carbon 14 its distribu tion in rain fresh waters and the oceans is ex tremely nonuniform This is inconvenient, since there is no standard value for its concentration, as in the case of carbon 14 A typical example of its application is in the study of geysers and hot springs These have been thought to consist in part of juvenile water that is water liberated from hot strata deep in the earth and reaching the surface for the first time Such water would contain no tritium The great bulk of water released by gey sers and hot springs contains tritium and so must be surface water which has sunk rather quickly to depths where high temperature prevails. A similar problem with a practical twist is the study of wells in arid regions If the well water is free of tritium it must be very old and its use for any purpose on a large scale results in mining an irreplaceable re source Quantitative studies of the tritium content of major natural waters gives some insight into their origin and pattern of circulation and mixing. Those studies in which carbon 14 tritium and stable isotope rates are measured together are most valuable Significant amounts of tritium and C^{14} have been released in recent years as a result of bomb tests

Of the remaining species in the table the beryl lium isotopes are the most studied Beryllium 7 has been measured in rain water and in atmospheric dust Its half life is quite short, comparable with the time scale of meteorological events and its use in meteorology in combination with the rarer P^{32} P^{33}, and S^{35} has begun Because of the short half life, one might expect the higher intensity of the cosmic rays near the poles to result in a higher fallout of beryllium 7 there This is not the case be cause of a countereffect the tropopause (upper

selected geologic structures such as salt structures and deep basins Much research and development work is being devoted to these problems. *See* DECONTAMINATION (RADIOACTIVE CONTAMINANTS), NUCLEAR FUELS NUCLEAR FUELS REPROCESSING, MONITORING (IONIZING RADIATION), NUCLEAR POWER, RADIOACTIVITY, RADIOCHEMICAL LABORATORY [K Z M]

Bibliography H Etherington (ed) *Nuclear Engineering Handbook*, 1958, K Saddington and W L. Templeton *Disposal of Radioactive Waste*, 1959 R Stephenson *Introduction to Nuclear Engineering*, 2d ed 1958

Radioactivity

A phenomenon resulting from instability of the ...e nucleus in certain atoms whereby the *nu* ... rably ... with ... overy ... model radio...

...er by W Rontgen and marked the ... of ...ear physics The radiations resulting from the ...ear transitions were subsequently studied and ...d to consist of three distinct types These were ...gnated as α-rays (helium nuclei) β rays (nega... or positive electrons) and γ rays (high fre...ncy electromagnetic radiation)

...1919 *Lord Rutherford discovered* that atomic ...lei could be broken up by bombardment with ...ys to form other nuclei In 1934 Irene Curie ... her husband Joliot first demonstrated that ...h artificially created nuclei themselves disinte...te spontaneously this phenomenon is called ...ficial or induced radioactivity All *chemical* ...ments may be *rendered radioactive* and the ...ilability *of this* wide variety of radioactive ...topes stimulated their use in science and tech...ogy in an enormous number of applications For ...liscussion of some of the more important applica...ns and a listing of other articles, see RADIO...CTIVITY (APPLICATIONS) See *also* ALPHA RAYS ...TA RAYS GAMMA RAYS [K W P]

A particular radioactive transition may be delayed by less than a microsecond or by more than a billion years, but the existence of a measurable delay *or* lifetime distinguishes a radioactive nuclear *transition from a so-called prompt nuclear* transition, such as is involved in the emission of most γ rays The delay is expressed quantitatively by the radioactive decay constant or by the mean life, or by the half period for each type of radioactive *atom For example, the half period of radium* is 1620 years, which means that on the average one-half of any initial supply of radium atoms will wait more than 1620 years before decaying and one-half will have decayed in less than 1620 years.

Types of radioactivity. There are at least six *common types of radioactivity characterized* by the *particular* type of nuclear radiation which is emitted by the transforming parent nucleus In one of these types α radioactivity the parent nucleus spontaneously emits an α-ray, then the atomic number, or nuclear charge, of the decay product is 2 *units less than that of the parent*, and the nuclear mass of the product is 4 atomic mass units less than that of the parent because the emitted α particle carries away this amount of nuclear charge and mass This decrease of two units of *atomic* number or nuclear charge between parent and *product means that the decay product will be a dif*...

... in column II of the periodic table. Its *decay product* after the emission of *an α ray* is a different chemical *element* radon, whose atomic number is 86 *and* whose position is in column 0 of the periodic table (*see* PERIODIC TABLE) The shift or displacement of the atomic number in radioactive decay is known as Soddy s displacement law The six types of radioactivity and the *displacement ΔZ* in atomic number which they produce are summarized in Table 1

TRANSITION RATES AND DECAY LAWS

Radioactive decay constant The rate of radioactive transformation or the activity of a source

Table 1 Types of radioactivity

Type	Symbol	Particles emitted	Change in atomic number ΔZ	Change in atomic mass number ΔA	Example
Alpha	α	Helium nucleus	−2	−4	$_{88}Ra^{226} \to {}_{86}Rn^{222}$
Beta negatron	β	Negative electron	+1	0	$_{11}Na^{24} \to {}_{12}Mg^{24}$
Beta positron	β+	Positive electron	−1	0	$_{11}Na^{22} \to {}_{10}Ne^{22}$
Electron capture	EC	Neutrino	−1	0	$_{4}Be^{7} \to {}_{3}Li^{7}$
Isomeric transition	IT	γ Rays and conversion electrons	0	0	$_{56}Ba^{137m} \to {}_{56}Ba^{137}$
Spontaneous fission	f	Heavy fragments	Various	Various	$_{92}U^{238} \to {}_{50}Sn^{132} + {}_{42}Mo^{106}$

m Isomeric (excited) state

preventive measures against corrosion and deterioration of the tanks or leakage of the wastes Tank storage is not considered to be permanent disposal but must be used until feasible methods of waste treatment and ultimate disposal are developed Up to January 1, 1957, in the United States, 62,000,000 gal of high level liquid wastes were in storage in 170 tanks The cost of waste storage in tanks has ranged from about 40 cents to $2 per gal

Liquid wastes of intermediate levels of activity from various chemical processes or relatively large experimental projects are of greater volume than high level wastes They may contain as much as 1/10 curie or more of radioactivity per gallon and are often high in dissolved chemical content Such wastes must be shielded to prevent external radiation and are not suitable for release to the general environment without extremely effective treatment for removal of the radioactive components When the radioactive components are removed, they must be concentrated and stored as in the case of high level waste In some locations intermediate level liquid wastes have been disposed of by dispersion in shallow soil formations where most of the radioactive elements are absorbed and retained for long periods by the soil materials

Shallow ground disposal is a subject of extensive study to determine the capacity of various soils in controlling the radioactivity of different waste solutions, particularly the more hazardous radionuclides which must not be dispersed to ground and surface waters A waste disposal system of this type has been used at the Oak Ridge National Laboratory since 1952 Through 1958, some 9,000,000 gal of liquid wastes containing about 115,000 curies of activity had been discharged into the seepage pit system The major operation of this kind has been at Hanford Wash where an unusual situation exists because of the soil formation and the isolated location, and where much larger quantities of liquid and radioactive materials are disposed of in shallow soil formations

Low level liquid wastes are present in large volumes of waste water from laboratory areas, decontamination operations, water used in basins to shield operators during work on radioactive materials, and other slightly contaminated liquids In favorable situations, where there are large volumes of surface water in isolated areas, such wastes are diluted and dispersed untreated or following partial decontamination by waste treatment processes At the Oak Ridge National Laboratory, for example, the contaminated waste water volume has been about 700 000 gal/day, of which roughly half has been discharged without treatment The remainder has been adequately decontaminated in a treatment plant employing the lime soda water softening process

Solid wastes. Solid radioactive wastes include such materials as machine turnings, nonusable contaminated equipment, and contaminated trash The activity may vary from a few times the background level to levels requiring shielding or remote handling In general, the disposal has been by land burial in selected areas where the soil has a capacity for retaining the radionuclides and the danger of excessive contamination of ground water is minimal The potential hazards and the care required in selecting, operating, and monitoring solid waste burial grounds depend, of course, on the particular radionuclides present as well as on the level of activity To a limited extent, solid wastes and concentrated low level liquid wastes have been packaged and disposed of by dumping at selected places in the ocean

Gaseous wastes. Gaseous radioactive wastes originate from such diverse places as air-cooled reactors, chemical processing plants laboratory hoods, and fissionable material fabrication facilities The levels of radioactivity vary with the type of operation, and the pollutants may be either gaseous or in the form of particles Deep-bed sand and fiber filters have been developed for the removal of particulate contaminants Equipment for absorbing iodine and other reactive gases is available, and inert gases can be removed by adsorption on charcoal or silica gel Disposal is usually by discharge into the atmosphere through tall stacks which provide dilution Continuous air monitors are used to determine the suitability of gaseous wastes for discharge under the particular conditions and to check the levels of air contamination that result after dispersion in the atmosphere

Future wastes from industry. Among a number of predictions concerning the growth of nuclear power in the United States summarized in AEC report WASH 742, J A Lane's estimate for the year 2000 is 7×10^8 kilowatts of heat.

From this and other predictions, it has been estimated that by the year 2000 high level liquid wastes will be produced at the rate of 50 000-500,000 gal/day By that time, the total accumulated volume of high level wastes will be of the order of 0 5 to 3×10^9 gal and the total accumulated radioactivity will be more than 10^{11} curies Because a considerable part of this accumulated activity will be due to strontium 90 and other long life radionuclides methods for ultimate disposal of these wastes must provide containment and control for at least several hundred years In a nuclear power industry large volumes of low and intermediate level wastes also will be produced Although the accumulated activity at any time will be orders of magnitude less than in the high level wastes, safe and economical disposal of these less-concentrated wastes must be achieved

The safe control of wastes from a nuclear power industry will involve a complex scheme of waste treatment, handling, and disposal Treatment may serve to remove the more hazardous radionuclides or to prepare the waste for disposal Temporary storage and transportation of the wastes may be necessary followed by one or more methods of ultimate disposal Prospective methods for ultimate disposal include (1) the fixation of the hazardous material in a stable solid medium and subsequent permanent storage or burial of the stable solid in selected locations, and [illegible] direct d[illegible]to

. . . The time interval over which the

Eq (3),

$$-\ln(A/A_0) = \ln(A_0/A) = \ln 2 = 0.693 = \lambda T \quad (8)$$

Then the half period T is related to the total radioactive decay constant λ, and to the mean life τ, by

$$T = 0.693/\lambda = 0.693\tau \quad (9)$$

For mnemonic reasons the half period T is much more frequently employed than the total decay constant λ or the mean life τ. For example, it is more common to speak of Th^{232} as having a half period of 1.4×10^{10} years than to speak of its mean life of 2.0×10^{10} years or its total decay constant of 1.6×10^{-18} sec^{-1}, although all three are equivalent statements of the average longevity of thorium 232 atoms.

The relationships between T, τ, and λ are summarized graphically in Fig. 1. Any initial activity $A_0\lambda$ is reduced to ½ in 1 half period T, to $1/e$ in 1 mean life τ, to ¼ in 2 half periods $2T$, etc. The slope of the activity curve, or rate of decrease of activity, is $d(A\lambda)/dt = -\lambda dA/dt = -\lambda(A\lambda)$. Thus the initial slope is $-\lambda(A_0\lambda) = -(A_0\lambda)/\tau$. The area under the activity curve, if integrated to $t = \infty$, is simply A_0, the total initial number of radioactive atoms. Also the initial activity $A_0\lambda$, if it could continue at a constant value for one mean life τ, would exactly destroy all the atoms because $(A_0\lambda)\tau = A_0$.

RADIOACTIVE SERIES DECAY

In a number of cases a radioactive nuclide A decays into a nuclide B which is also radioactive; the nuclide B decays into C which is also radioactive, etc. For example, $_{90}Th^{232}$ decays into a long series of 10 successively radioactive nuclides, as described later in the discussion of radioactive transformation series. Substantially all the primary products of nuclear fission are negatron β ray emitters which decay through a chain or series of 2–6 successive β ray emitters before a stable nuclide is reached as an end product. See FISSION, NUCLEAR.

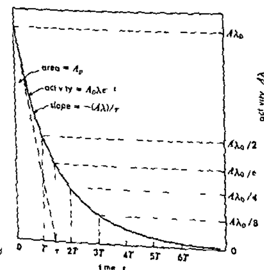


Fig. 1. Graphical relationships in decay of a single radioactive nuclide.

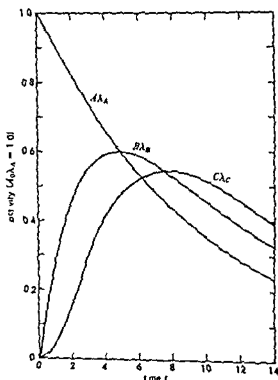


Fig. 2. Growth and decay of activity $B\lambda_B$ of daughter product, and $C\lambda_C$ of granddaughter product, in an initially pure source of a radioactive parent whose activity at $t = 0$ is $A_0\lambda_A$.

Let the initial part of such a series be represented by



$$A \xrightarrow{\lambda_A} B \xrightarrow{\lambda_B} C \xrightarrow{\lambda_C} D \xrightarrow{\lambda_D}$$

where radioactive atoms of types $A, B, C, D, \ldots$ have radioactive decay constants given by $\lambda_A, \lambda_B, \lambda_C, \lambda_D, \ldots$. Then if there are initially present, at time $t = 0$, only A_0 atoms of type A, the numbers $A, B, C, \ldots$, of atoms of types $A, B, C, \ldots$, which will be present at a later time t are given by

$$A = A_0 e^{-\lambda_A t} \quad (10)$$

$$B = A_0 \frac{\lambda_A}{\lambda_B - \lambda_A}(e^{-\lambda_A t} - e^{-\lambda_B t}) \quad (11)$$

$$C = A_0 \left(\frac{\lambda_A}{\lambda_C - \lambda_A} \frac{\lambda_B}{\lambda_B - \lambda_A} e^{-\lambda_A t} + \frac{\lambda_A}{\lambda_A - \lambda_B} \frac{\lambda_B}{\lambda_C - \lambda_B} e^{-\lambda_B t} + \frac{\lambda_A}{\lambda_A - \lambda_C} \frac{\lambda_B}{\lambda_B - \lambda_C} e^{-\lambda_C t} \right) \quad (12)$$

and the activities of $A, B, C, \ldots$ are $A\lambda_A, B\lambda_B, C\lambda_C, \ldots$. General equations describing the

equals the number A of identical radioactive atoms present in the source multiplied by their characteristic radioactive decay constant λ. Thus

$$\text{Activity} = A\lambda \quad \text{disintegrations per second} \qquad (1)$$

where the decay constant λ has dimensions of sec^{-1}. The numerical value of λ expresses the statistical probability of decay of each radioactive atom in a group of identical atoms per unit time. For example, if $\lambda = 0.01 \ sec^{-1}$ for a particular radioactive species, then each atom has a chance of 0.01 (1%) of decaying in 1 sec and a chance of 0.99 (99%) of not decaying in any given 1-sec interval. The constant λ is one of the most important characteristics of each radioactive nuclide. λ is essentially independent of all physical and chemical conditions such as temperature, pressure, concentration, chemical combination, or age of the radioactive atoms. The half period, to be discussed later, is inversely proportional to λ.

Among the more than 800 known radioactive nuclides, no two have exactly the same decay constant. The identification of some radioactive samples can be made simply by measuring λ, which then serves as an equivalent of qualitative chemical analysis. For the most common radioactive nuclides, the range of λ extends from $3 \times 10^{6} \ sec^{-1}$ (for thorium C′) to $1.6 \times 10^{-18} \ sec^{-1}$ (for thorium).

Dual decay. Many radioactive nuclides have two or more independent and alternative modes of decay. For example, U^{238} (uranium-238) can decay either by α-ray emission or by spontaneous fission. A single atom of Cu^{64} (copper-64) can decay in any of three competing independent ways: negatron β-ray emission, positron β-ray emission, or electron capture. When two or more independent modes of decay are possible, the nuclide is said to exhibit dual decay.

The competing modes of decay of any nuclide have independent partial decay constants given by the probabilities $\lambda_1, \lambda_2, \lambda_3, \ldots$ per second, and the total probability of decay is represented by the total decay constant λ, where

$$\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \cdots \qquad (2)$$

If there are A identical atoms present, the partial activities as measured by the different modes of decay are $A\lambda_1, A\lambda_2, A\lambda_3, \ldots$ and the total activity $A\lambda$ is

$$A\lambda = A\lambda_1 + A\lambda_2 + A\lambda_3 + \cdots \qquad (2a)$$

The partial activities, $A\lambda_1, \ldots$, such as positron β rays from Cu^{64}, are proportional to the total activity $A\lambda$ at all times.

The branching ratio is the fraction of the decaying atoms which follow a particular mode of decay and equals $A\lambda_1/A\lambda$ or λ_1/λ. For example, in the case of Cu^{64} the measured branching ratios are $\lambda_1/\lambda = 0.40$ for negatron β-decay, $\lambda_2/\lambda = 0.20$ for positron β-decay, and $\lambda_3/\lambda = 0.40$ for electron capture. The sum of all the branching ratios for a particular nuclide is unity.

Exponential decay law. The total activity $A\lambda$ equals the rate of decrease $-dA/dt$ in the number of radioactive atoms A present. Because λ is independent of the age t of an atom, integration of the differential equation of radioactive decay $-dA/dt = A\lambda$ gives

$$\ln \frac{A}{A_0} = -\lambda(t - t_0) \qquad (3)$$

where ln represents the natural logarithm to the base e, and A atoms remain at time t if there were A_0 atoms initially present at time t_0. If $t_0 = 0$, then Eq. (3) can be rewritten as the exponential law of radioactive decay in its most common form

$$A = A_0 e^{-\lambda t} \qquad (4)$$

The initial activity at $t = 0$ was $A_0\lambda$, and the activity at t, when only A atoms remain untransformed, is $A\lambda$. Because λ is a constant, the fractional activity $A\lambda/A_0\lambda$ at time t and the fractional amount of radioactive atoms A/A_0 are given by

$$\frac{A\lambda}{A_0\lambda} = \frac{A}{A_0} = e^{-\lambda t} \qquad (5)$$

Note that in cases of dual decay the partial activities $A\lambda_1, A\lambda_2, \ldots$, also decrease with time as $e^{-\lambda t}$, not as $e^{-\lambda_1 t}, \ldots$, because $A\lambda_1/A_0\lambda_1 = A/A_0 = e^{-\lambda t}$, where λ is the total decay constant. This is because the decrease of each partial activity with time is due to the depletion of the total stock of atoms A, and this depletion is accomplished by the combined action of all the competing modes of decay.

Mean life. The actual life of any particular atom can have any value between zero and infinity. The average or mean life of a large number of identical radioactive atoms is, however, a definite and important quantity.

If there are A_0 atoms present initially at $t = 0$, then the number remaining undecayed at a later time t is $A = A_0 e^{-\lambda t}$ by Eq. (4). Each of these A atoms has a life longer than t. In an additional infinitesimally short time interval dt between time t and $t + dt$, the absolute number of atoms which will decay on the average is $A\lambda \, dt$, and these atoms had a life span t. The total L of the life spans of all the A_0 atoms is the sum or integral of $tA\lambda \, dt$ from $t = 0$ to $t = \infty$, which is

$$L = \int_0^\infty tA\lambda \, dt = \int_0^\infty tA_0\lambda e^{-\lambda t} \, dt = \frac{A_0}{\lambda} \qquad (6)$$

Then the average lifetime L/A_0, which is called the mean life τ, is simply

$$\tau = 1/\lambda \qquad (7)$$

where λ is the *total* radioactive decay constant of Eq. (2). Substitution of $t = \tau = 1/\lambda$ into Eq. (5) shows that the mean life is the time required for the number of atoms, or their activity, to fall to $e^{-1} = 0.368$ of any initial value.

Half period. The time interval over which the chance of survival of a particular radioactive atom is exactly one-half is called the half period T. From Eq (3),

$$-\ln(A/A_0) = \ln(A_0/A) = \ln 2 = 0.693 = \lambda T \quad (8)$$

Then the half period T is related to the total radioactive decay constant λ, and to the mean life τ, by

$$T = 0.693/\lambda = 0.693\tau \quad (9)$$

For mnemonic reasons, the half period T is much more frequently employed than the total decay constant λ or the mean life τ. For example, it is more ... to speak of Th^{232} as having a half period

...

atoms.

The relationships between T, τ, and λ are summarized graphically in Fig. 1. Any initial activity $A_0\lambda$ is reduced to ½ in 1 half period T, to $1/e$ in 1 mean life τ, to ¼ in 2 half periods $2T$, etc. The slope of the activity curve, or rate of decrease of activity, is $d(A\lambda)/dt = -\lambda dA/dt = -\lambda(A\lambda)$. Thus the initial slope is $-\lambda(A_0\lambda) = -(A_0\lambda)/\tau$. The area under the activity curve, if integrated to $t = \infty$, is simply A_0, the total initial number of radioactive atoms. Also the initial activity $A_0\lambda$, if it could continue at a constant value for one mean life τ, would exactly destroy all the atoms because $(A_0\lambda)\tau = A_0$.

RADIOACTIVE SERIES DECAY

In a number of cases a radioactive nuclide A decays into a nuclide B which is also radioactive, the nuclide B decays into C which is also radioactive, etc. For example, $_{90}Th^{232}$ decays into a long series of 10 successively radioactive nuclides as de-

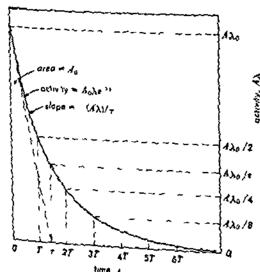


Fig 1 Graphical relationships in decay of a single radioactive nuclide

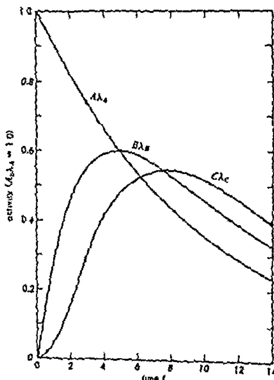
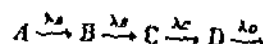


Fig 2 Growth and decay of activity $B\lambda_B$ of daughter product, and $C\lambda_C$ of granddaughter product, in an initially pure source of a radioactive parent whose activity at $t = 0$ is $A_0\lambda_A$

scribed later in the discussion of radioactive transformation series. Substantially all the primary products of nuclear fission are negatron β ray emitters which decay through a chain or series of 2-6 successive β ray emitters before a stable nuclide is reached as an end product. See FISSION, NUCLEAR.

Let the initial part of such a series be represented by



$$A \xrightarrow{\lambda_A} B \xrightarrow{\lambda_B} C \xrightarrow{\lambda_C} D \xrightarrow{\lambda_D}$$

where radioactive atoms of types $A, B, C, D, \ldots$, have radioactive decay constants given by $\lambda_A, \lambda_B, \lambda_C, \lambda_D, \ldots$. Then if there are initially present, at time $t = 0$, only A_0 atoms of type A, the numbers $A, B, C, \ldots$, of atoms of types $A, B, C, \ldots$, which will be present at a later time t are given by

$$A = A_0 e^{-\lambda_A t} \quad (10)$$

$$B = A_0 \frac{\lambda_A}{\lambda_B - \lambda_A}(e^{-\lambda_A t} - e^{-\lambda_B t}) \quad (11)$$

$$C = A_0 \left(\frac{\lambda_A}{\lambda_C - \lambda_A} \frac{\lambda_B}{\lambda_B - \lambda_A} e^{-\lambda_A t} + \frac{\lambda_A}{\lambda_A - \lambda_B} \frac{\lambda_B}{\lambda_C - \lambda_B} e^{-\lambda_B t} + \frac{\lambda_A}{\lambda_A - \lambda_C} \frac{\lambda_B}{\lambda_B - \lambda_C} e^{-\lambda_C t} \right) \quad (12)$$

and the activities of $A, B, C, \ldots$, are $A\lambda_A$, $B\lambda_B$, $C\lambda_C, \ldots$. General equations describing the

amounts and activities of any number of radioactive decay products are more complicated and are given in standard texts.

Figure 2 illustrates the growth and decay of the activity of a short series of radioactive decay products in accord with Eqs. (10) to (12).

Radioactive equilibrium. In Fig. 2 note that the ratio $B\lambda_B/A\lambda_A$ of the activities of the parent A and the daughter product B change with time. The activity $B\lambda_B$ is zero initially and also after a very long time, when all the atoms have decayed. Thus $B\lambda_B$ passes through a maximum value, and it can be shown that this occurs at a time t_m given by

$$t_m = \frac{\ln(\lambda_B/\lambda_A)}{(\lambda_B - \lambda_A)} \qquad (13)$$

The situation in which the activities $A\lambda_A$ and $B\lambda_B$ are exactly equal to each other is called ideal equilibrium, and exists only at the moment t_m.

If the parent A is longer lived than the daughter B, as occurs in many cases, then at a time which is long compared with the mean life τ_B of B, the activity ratio approaches a constant value given by

$$\frac{B\lambda_B}{A\lambda_A} = \frac{\lambda_B}{\lambda_B - \lambda_A} = \frac{T_A}{T_A - T_B} \qquad (14)$$

where T_A and T_B are the half periods of A and of B. When the activity ratio $B\lambda_B/A\lambda_A$ is constant, a particular type of "radioactive equilibrium" exists. This is spoken of as secular equilibrium if the activity ratio is experimentally indistinguishable from unity, as occurs when T_A is very much greater than T_B.

Equilibrium concepts are applied also between a long lived parent and any of its decay products in a long series. For example, in a sufficiently old uranium ore, radium ($T = 1620$ years) is in secular equilibrium with its ultimate parent uranium ($T = 4.5 \times 10^9$ years) although there are four intermediate radioactive substances intervening in the series between uranium and radium. Here, secular equilibrium expresses the fact that the activities of radium and uranium continue to be equal to each other even though the activity of the parent uranium is decreasing with time.

When T_B is comparable with T_A, Eq. (14) shows that the equilibrium ratio will clearly exceed unity; this situation is spoken of as transient equilibrium. For example, in the fission product decay series



$$Ba^{140} \rightarrow La^{140} \rightarrow Ce^{140}$$

the half period of Ba^{140} is 307 hours, and that of La^{140} is 40 hours. In an initially pure source of Ba^{140} the activity of La^{140} starts at zero, rises to a maximum at $t_m = 135$ hours (Eq. 13), then decreases, and after a few hundred hours is in transient equilibrium with its parent, when the La^{140} activity (by Eq. 14) is $307/(307-40) = 1.15$ times the activity of its parent Ba^{140}.

Natural radioactivity. The radioactivity possessed by a few naturally occurring isotopes without the intervention of man or machines is called natural radioactivity. Because the earth is composed of atoms which were believed to have been created more than 3×10^9 years ago, the naturally occurring parent radioactive isotopes are those which have such long half periods that detectable residual activity is still observable today. For example, present day uranium is an isotopic mixture containing 99.3% U^{238} whose half period is 4.5×10^9 years, and 0.7% of the shorter lived uranium isotope U^{235} whose half period is 0.7×10^9 years. Geophysical evidence indicates that originally some U^{236} was present also but none is found in nature now because its half period of 0.02×10^9 years is less than 1/100 of the age of the earth and therefore substantially total decay has occurred.

Besides uranium, the most important naturally occurring parent radioactive element is thorium, whose single long lived isotope Th^{232}, has a half period of 14×10^9 years, or more than 3 times the assumed age of the earth.

Uranium 238 decays through a long series of 14 radioactive decay products involving 8 cases of α radioactivity and 6 cases of β radioactivity be-

Table 2. Parent radioactive nuclides which are found in nature

Nuclide: Atomic number Z	Nuclide: Mass number A	Atom % abundance in element	Half period, years	Radioactive transitions observed	Disintegration energy, Mev
19 K	40	0.0119	1.2×10^9	β, LC, γ	1.4
37 Rb	87	27.85	6×10^{10}	β	0.3
49 In	115	95.77	6×10^{14}	β	0.6
52 Te	130	34.49	1×10^{21}	Growth of $_{54}Xe^{130}$	1.6
57 La	138	0.089	2×10^{11}	β, LC	3
60 Nd	144	23.9	1×10^{15}	α	1.9
62 Sm	147	15.07	1.4×10^{11}	α	2.1
71 Lu	176	2.6	7.5×10^{10}	β, γ	0.9
75 Re	187	62.93	4×10^{12}	β	0.4
90 Th	232	100	1.4×10^{10}	α	4.05
92 U	235	0.715	7.1×10^8	α	4.66
92 U	238	99.28	4.5×10^9	α	4.25

ending as a stable isotope of lead, Pb^{206} of these members of the U^{238} decay chain very short half periods so their existence in re is entirely dependent on the presence of long lived parent, and thus is a genealogical ... in nature ... The de... ecay se... series of ... of Th^{232} supports 10, short lived but naturally ioactive substances

... mmon elements contain long lived

e isotope K^{40} which has a half period ... years and emits negatron β rays and γ rays in dual decay to stable Ca^{40} and A^{40} This isotope the principal source of natural radioactivity in the normal human being each human contains bout 0 1 microcurie of the naturally radioactive otassium isotope K^{40} A number of laboratories f physics and radiobiology have sensitive radiation ... In one rate

K^{40} content. *See* MONITORING (IONIZING RADIATION)

Table 2 summarizes the radioactive properties of all the well established cases of natural radioactivity Geological age measurements are based on the accumulation of decay products of these long lived isotopes especially in the cases of K^{40} Rb^{87} Th^{232} U^{235} and U^{238} See GEOCHRONOMETRY

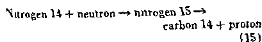
Induced radioactivity This is the property of radioactivity exhibited by newly created atoms wh h are prod iced by nuclear reactions It is

duced or induced in the laboratory as the product ... ion ... of ... pro- duce compound nuclei of nitrogen 15 which promptly emit a proton (hydrogen nucleus) leav

ple by the nuclear reaction

$$\text{Nitrogen 14} + \text{neutron} \rightarrow \text{nitrogen 15} \rightarrow \text{carbon 14} + \text{proton} \quad (15)$$

This reaction is easily carried out using many types of nuclear machines such as the cyclotron Van de Graaff Generator linear accelerator or in any nuclear reactor because each of these is a source of the neutrons which are needed for producing or inducing this particular nuclear transmutation This particular transmutation reaction is one which happens to occur in nature also without the intervention of man or machines because the nitrogen in the earth s atmosphere is continually bombarded by neutrons which enter the atmosphere as cosmic rays, thus producing induced radioactivity as newly formed atoms of carbon 14 *See* COSMIC RAYS, NUCLEAR REACTION, PARTICLE ACCELERATOR, REACTOR, NUCLEAR, TERRESTRIAL NUCLEAR REACTIONS



As a result of extensive nuclear transmutation and disintegration experiments using the 274 known stable nuclides as well as uranium and thorium as targets more than 800 new nuclides have been produced and studied Each of these is radioactive and is therefore an instance of induced radioactivity

Tables of the identity and properties of these radioactive nuclides, or isotopes, are being revised constantly to include newly discovered information The most up to date tables are routinely published

a target material A is bombarded to produce a radioactive product B whose radioactive decay constant is λ_B the number of atoms B which are present after a bombardment of duration t, and their activity $B\lambda_B$, are given by

$$B\lambda_B = \left(\frac{Y}{\lambda_B}\right)(1 - e^{-\lambda_B t}) \quad (16)$$

where the yield Y, has dimensions equivalent to curies of activity produced per second of bombardment The yield depends on the number of atoms A present in the target, the intensity of the beam

cle under the condi ...

Radioactive transformation series. As noted in Eqs (10) to (12) many radioactive substances have decay products which are also radioactive Thus many long chains or series of radioactive transformations are known

The three naturally occurring transformation series are headed by Th^{232}, U^{235}, and U^{238} Their genealogical relationships are summarized in Figs 3-5

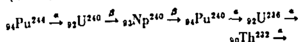
Each of the naturally occurring radioactive isotopes in these transformation series has two synonymous names For example, the commercially important radioisotope whose classical name is mesothorium 1 is isotopic with radium 226, but has a mass number of 228, mesothorium 1 whose symbol is $MsTh_1$ can therefore be designated also as radium 228 (symbol Ra^{228}) Table 3 summarizes the names symbols and some radioactive properties of the transformation series whose family relationships are shown in Figs 3-5

A number of induced activities which are also members of these three transformation series have been produced in transmutation experiments For example plutonium 244 (Pu^{244}) has a half period of 76,000 years and decays by α emission into ura

Table 3. Names, symbols, and radioactive properties of members of the three naturally occurring radioactive transformation series

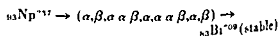
Conventional name	Conventional symbol	Atomic number	Mass number	Isotopic symbol	Half-period	Type of decay
Uranium (4*n* + 2) series						
Uranium I	UI	92	238	U^{238}	4.5×10^9y	α
Uranium X_1	UX_1	90	234	Th^{234}	24 d	β^-
Uranium X_2	UX_2	91	234	Pa^{234m}	1.2m	IT, β
Uranium Z	UZ	91	234	Pa^{234}	6.7h	β^-
Uranium II	UII	92	234	U^{234}	2.5×10^5y	α
Ionium	Io	90	230	Th^{230}	8×10^4y	α
Radium	Ra	88	226	Ra^{226}	1620y	α
Radon	Rn	86	222	Rn^{222}	3.8d	α
Radium A	RaA	84	218	Po^{218}	3.0m	α
Radium B	RaB	82	214	Pb^{214}	27 m	β^-
Radium C	RaC	83	214	Bi^{214}	20 m	β^-, α
Radium C′	RaC′	84	214	Po^{214}	1.6×10^{-4}s	α
Radium C″	RaC″	81	210	Tl^{210}	1.3m	β^-
Radium D	RaD	82	210	Pb^{210}	22 y	β^-
Radium E	RaE	83	210	Bi^{210}	5.0d	β^-
Radium F	RaF	84	210	Po^{210}	138 d	α
Polonium	Po	84	210	Po^{210}	138 d	α
Radium G	RaG	82	206	Pb^{206}	Stable	Stable
Thorium (4*n*) series						
Thorium	Th	90	232	Th^{232}	1.4×10^{10}y	α
Mesothorium$_1$	$MsTh_1$	88	228	Ra^{228}	6.7y	β^-
Mesothorium$_2$	$MsTh_2$	89	228	Ac^{228}	6.1h	β^-
Radiothorium	RdTh	90	228	Th^{228}	1.9y	α
Thorium X	ThX	88	224	Ra^{224}	3.6d	α
Thoron	Tn	86	220	Rn^{220}	54 s	α
Thorium A	ThA	84	216	Po^{216}	0.16s	α
Thorium B	ThB	82	212	Pb^{212}	10.6h	β^-
Thorium C	ThC	83	212	Bi^{212}	1.0h	β^-, α
Thorium C′	ThC′	84	212	Po^{212}	3×10^{-7}s	α
Thorium C″	ThC″	81	208	Tl^{208}	3.1m	β^-
Thorium D	ThD	82	208	Pb^{208}	Stable	Stable
Actinium (4*n* + 3) series						
Actinouranium	AcU	92	235	U^{235}	7.1×10^8y	α
Uranium Y	UY	90	231	Th^{231}	25 h	β^-
Protactinium	Pa	91	231	Pa^{231}	3.4×10^4y	α
Actinium	Ac	89	227	Ac^{227}	22 y	β^-, α
Radioactinium	RdAc	90	227	Th^{227}	18 d	α
Actinium K	AcK	87	223	Fr^{223}	22 m	β^-, α
Actinium X	AcX	88	223	Ra^{223}	11 d	α
Astatine	At	85	219	At^{219}	0.9m	α, β^-
Actinon	An	86	219	Rn^{219}	3.9s	α
Actinium A	AcA	84	215	Po^{215}	1.8×10^{-3}s	α
Actinium B	AcB	82	211	Pb^{211}	36 m	β^-
Actinium C	AcC	83	211	Bi^{211}	2.2m	α, β^-
Actinium C′	AcC′	84	211	Po^{211}	0.5s	α
Actinium C″	AcC″	81	207	Tl^{207}	4.8m	β^-
Actinium D	AcD	82	207	Pb^{207}	Stable	Stable

nium 240 (U^{240}), thence by two β transformations and two more α transformations into thorium 232 (Th^{232}), according to the following scheme:



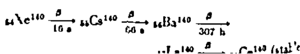
$$_{94}Pu^{244} \xrightarrow{\alpha} {}_{92}U^{240} \xrightarrow{\beta} {}_{93}Np^{240} \xrightarrow{\beta} {}_{94}Pu^{240} \xrightarrow{\alpha} {}_{92}U^{236} \xrightarrow{\alpha} {}_{90}Th^{232} \xrightarrow{\alpha}$$

The so-called neptunium series, or $4n + 1$ series, does not exist in nature because of the length of the half period (2 000 000 yr) of its longest lived member, $_{93}Np^{237}$. This series has been produced and studied in transmutation experiments, and follows the pattern.

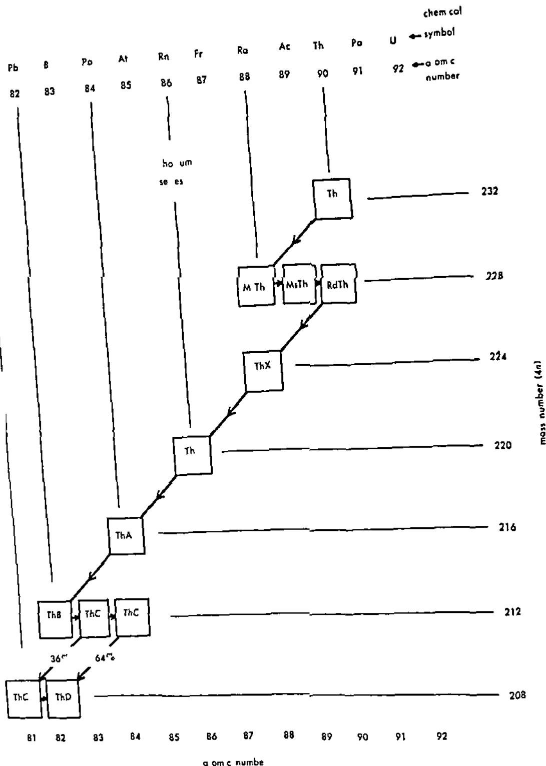


$$_{93}Np^{237} \rightarrow (\alpha,\beta,\alpha\ \alpha\ \beta,\alpha,\alpha\ \alpha\ \beta,\alpha,\beta) \rightarrow {}_{83}Bi^{209}\text{(stable)}$$

More than 90 transformation series have been studied among the radioactive products of nuclear fission. These series vary in length from 2 6 radioactive members, each of which emits negatron β rays. An important example is



$$_{54}Xe^{140} \xrightarrow[16\ s]{\beta} {}_{55}Cs^{140} \xrightarrow[66\ s]{\beta} {}_{56}Ba^{140} \xrightarrow[307\ h]{\beta} {}_{57}La^{140} \xrightarrow{\beta} {}_{58}Ce^{140}\text{ (stable)}$$

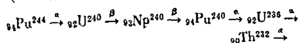


Fg 3 Ma n l ne of decay for the tho um series of heavy rad oact ve nucl des. D agonal a ows denote α-decay n wh ch atom c number dec eases by 2 un ts and mass number dec eases by 4 un ts. Ho zontal ar rows deno e nega ron β-decay n wh ch atom c number Inc eases by 1 un t and mass number s unchanged Every member of this ser es has a mass number g ven by 4n whe e n s an nteger Hence the tho um ser es s also called the 4n ser es Note dual-decay of ThC 36% by α-decay to ThC and 64% by β-decay to ThC

Table 3 Names, symbols, and radioactive properties of members of the three naturally occurring radioactive transformation series

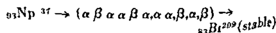
Conventional name	Conventional symbol	Atomic number	Mass number	Isotopic symbol	Half-period	Type of decay
Uranium $(4n+2)$ series						
Uranium I	UI	92	238	U^{238}	4.5×10^9y	α
Uranium X_1	UX_1	90	234	Th^{234}	24 d	β
Uranium X_2	UX_2	91	234	Pa^{234m}	1.2m	IT, β
Uranium Z	UZ	91	234	Pa^{234}	6.7h	β
Uranium II	UII	92	234	U^{234}	2.5×10^5y	α
Ionium	Io	90	230	Th^{230}	8×10^4y	α
Radium	Ra	88	226	Ra^{226}	1620y	α
Radon	Rn	86	222	Rn^{222}	3.8d	α
Radium A	RaA	84	218	Po^{218}	3.0m	α
Radium B	RaB	82	214	Pb^{214}	27 m	β
Radium C	RaC	83	214	Bi^{214}	20 m	β α
Radium C′	RaC′	84	214	Po^{214}	1.6×10^{-4}s	α
Radium C″	RaC″	81	210	Tl^{210}	1.3m	β^-
Radium D	RaD	82	210	Pb^{210}	22 y	β
Radium E	RaE	83	210	Bi^{210}	5.0d	β
Radium F	RaF	84	210	Po^{210}	138 d	α
Polonium	Po	84	210	Po^{210}	138 d	α
Radium G	RaG	82	206	Pb^{206}	Stable	Stable
Thorium $(4n)$ series						
Thorium	Th	90	232	Th^{232}	1.4×10^{10}y	α
Mesothorium$_1$	$MsTh_1$	88	228	Ra^{228}	6.7y	β^-
Mesothorium$_2$	$MsTh_2$	89	228	Ac^{228}	6.1h	β^-
Radiothorium	RdTh	90	228	Th^{228}	1.9y	α
Thorium X	ThX	88	224	Ra^{224}	3.6d	α
Thoron	Tn	86	220	Rn^{220}	54.5s	α
Thorium A	ThA	84	216	Po^{216}	0.16s	α
Thorium B	ThB	82	212	Pb^{212}	10.6h	β
Thorium C	ThC	83	212	Bi^{212}	1.0h	β^-, α
Thorium C′	ThC′	84	212	Po^{212}	3×10^{-7}s	α
Thorium C″	ThC″	81	208	Tl^{208}	3.1m	β^-
Thorium D	ThD	82	208	Pb^{208}	Stable	Stable
Actinium $(4n+3)$ series						
Actinouranium	AcU	92	235	U^{235}	7.1×10^8y	α
Uranium Y	UY	90	231	Th^{231}	25 h	β
Protactinium	Pa	91	231	Pa^{231}	3.4×10^4y	α
Actinium	Ac	89	227	Ac^{227}	22 y	β^-, α
Radioactinium	RdAc	90	227	Th^{227}	18 d	α
Actinium K	AcK	87	223	Fr^{223}	22 m	β^-, α
Actinium X	AcX	88	223	Ra^{223}	11 d	α
Astatine	At	85	219	At^{219}	0.9m	α β^-
Actinon	An	86	219	Rn^{219}	3.9s	α
Actinium A	AcA	84	215	Po^{215}	1.8×10^{-3}s	α
Actinium B	AcB	82	211	Pb^{211}	36 m	β
Actinium C	AcC	83	211	Bi^{211}	2.2m	α β
Actinium C	AcC	84	211	Po^{211}	0.5s	α
Actinium C	AcC′	81	207	Tl^{207}	4.8m	β^-
Actinium D	AcD	82	207	Pb^{207}	Stable	Stable

nium 240 (U^{240}), thence by two β transformations and two more α transformations into thorium 232 (Th^{232}), according to the following scheme



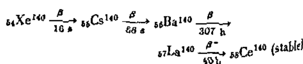
$$_{94}Pu^{244} \xrightarrow{\alpha} {}_{92}U^{240} \xrightarrow{\beta} {}_{93}Np^{240} \xrightarrow{\beta} {}_{94}Pu^{240} \xrightarrow{\alpha} {}_{92}U^{236} \xrightarrow{\alpha} {}_{90}Th^{232} \xrightarrow{\alpha}$$

The so called neptunium series or $4n + 1$ series does not exist in nature because of the length of the half period (2,000 000 yr) of its longest lived member, $_{93}Np^{237}$. This series has been produced and studied in transmutation experiments, and follows the pattern



$$_{93}Np^{237} \rightarrow (\alpha\, \beta\, \alpha\, \alpha\, \beta\, \alpha, \alpha\, \alpha, \beta, \alpha, \beta) \rightarrow {}_{83}Bi^{209} (stable)$$

More than 90 transformation series have been studied among the radioactive products of nuclear fission These series vary in length from 2 6 radioactive members, each of which emits negatron β rays An important example is



$$_{54}Xe^{140} \xrightarrow[16\,s]{\beta} {}_{55}Cs^{140} \xrightarrow[66\,s]{\beta} {}_{56}Ba^{140} \xrightarrow[307\,h]{\beta} {}_{57}La^{140} \xrightarrow[40\,h]{\beta^-} {}_{58}Ce^{140} \text{ (stable)}$$

ps, or lines, in the spectrum of α rays from articular radionuclide For example, in the cay of a large group of U^{238} atoms, 77% of α decays will be by emission of α rays whose etic energy is 4 20 Mev (million electron volts), le 23% will be by emission of 4 15-Mev α rays en the 4 20-Mev α ray is emitted the decay prod-

uct nucleus is formed in its ground (lowest en ... When a 4 15 Mev α ray is emitted, the ...

0 05 Mev γ ray or alternatively by the ... the same amount of energy in the form of a con

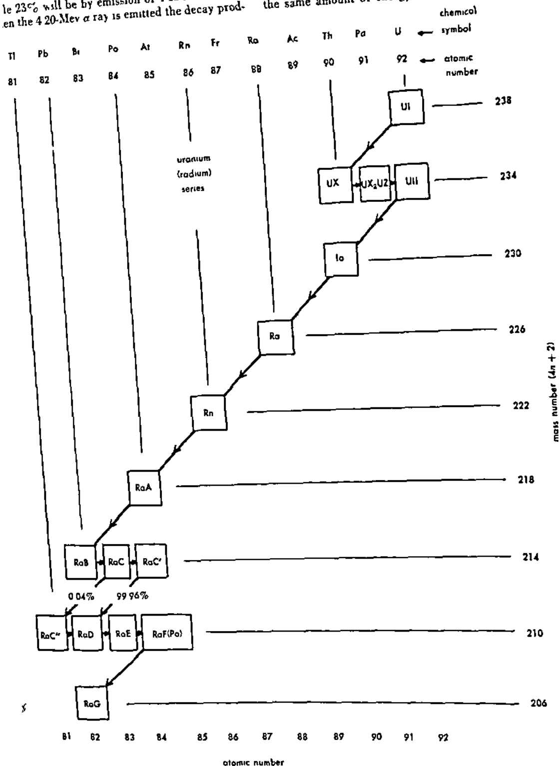


Fig 5 Main line of decay of uranium series, or 4n + 2 series, of heavy radioactive nuclides, headed in nature by uranium-238 Each member has a mass number given by 4n + 2, where n is an integer

where the half period of each negatron β transformation is given under the corresponding arrow (s = seconds h = hours)

ALPHA-RAY DECAY

Alpha ray decay is that type of radioactivity in which the parent nucleus expels an α ray The α ray is emitted with a speed of the order of 1 to 2 X 10^9 cm/sec, that is about 1/20 of the velocity of light

In the simplest case of α decay every α ray would be emitted with exactly the same velocity and hence the same kinetic energy However in most cases there are two or more discrete energy

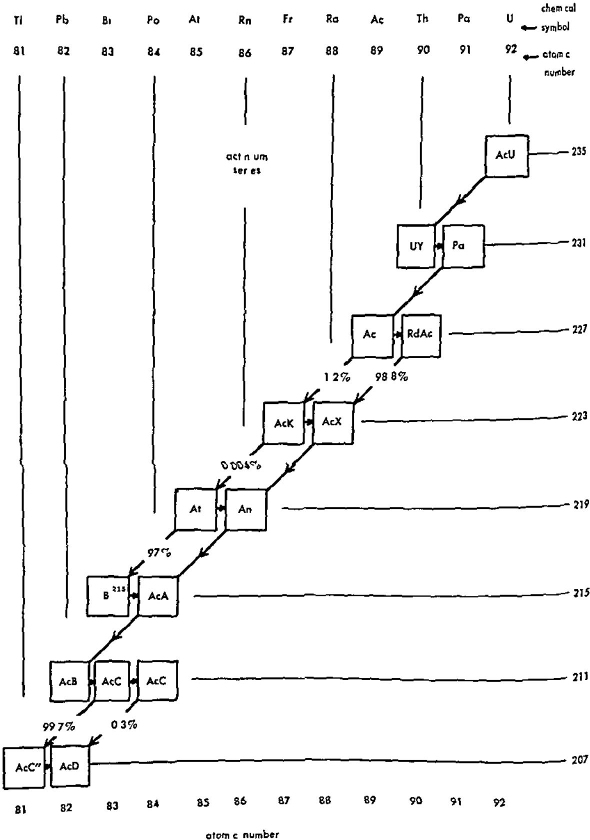


Fig 4 Main line of decay of actinium series or $4n + 3$ series of heavy radioactive nuclides headed in nature by uranium 235 Each member has a mass number given by $4n + 3$ where n is an integer

;h some light does get through, as in the case e thin gold leaf which is sometimes used for ·ing signs on store windows

ie wave-mechanical probability of the trans- ion of an α particle through the nuclear poten barrier is very strongly dependent upon the ·gy of the emitted α ray Analytically the prob- ity of transmission T depends exponentially n a barrier transmission exponent γ accord to

$$T = e^{-\gamma} \tag{18}$$

a good approximation

$$= \left(\frac{4\pi^2}{h}\right)\frac{(Z-2)2e^2}{V} - \left(\frac{8\pi}{h}\right)[2(Z-2)2e^2MR]^{1/2} \tag{19}$$

here $h = 6\,625 \times 10^{-2}$ erg second is Planck's onstant and M is the so-called reduced mass of he α particle For the α-decay of Ra^{226}, the nu nerical value of γ is about 71 hence $T = e^{-1} =$ 10^{31} The first term on the right side of Eq (19) is about 154 and is therefore the dominant term When this term is taken alone,

$$e^{-(4\pi^2/h)\,(Z-2)2e^2/V}$$

is called the Gamow factor for barrier penetra- tion

Inspection of Eq (19) shows that the barrier transmission decreases with increasing nuclear charge $(Z-2)e$, increases with increasing veloc ity V of emission of the α ray, and increases with increasing radius R of the nucleus When the ex perimentally known values of α decay energy are substituted into Eq (19), with R about 10^{12} cm and Z about 90, the transmission coefficient $T = e^{-\gamma}$ is found to extend over a domain of about $10^{\,0}$ ... of about 10^{0} is just what is

t

[illegible]

tion (19) thus explains the Geiger Nuttall ... very successfully Figure 8 presents a logical modern form of the Geiger Nuttall relationship, as used extensively by I Perlman and coworkers The

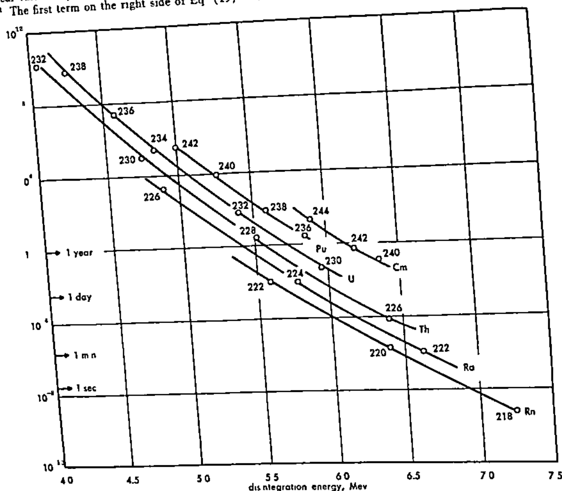


Fig 8 Systematics of the broad range of half periods for α ray decay and their strong dependence on α-decay energy and weaker dependence on nuclear charge Numbers beside experimental points are mass numbers of parent α-ray emitters Lines connect parent isotopes and are drawn using wave mechanical theory of α ray transmission through nuclear potential barriers

version electron and the associated spectrum of characteristic x rays Thus all α ray spectra are line spectra, with α rays emitted in one or more discrete and homogeneous energy groups, and α ray spectra are accompanied by a γ ray spectrum when ever there are two or more α ray groups in the spectrum

Geiger Nuttall rule. Among all the known α ray emitters most α ray energy spectra lie in the do main of 4–6 Mev, although a few extend as low as 2 Mev ($_{62}Sm^{147}$) and as high as 10 Mev (ThC′) There is a systematic relationship between the ki netic energy of the emitted α rays and the half period of the α emitter The highest energy α rays are emitted by short lived nuclides and the lowest energy α rays are emitted by the very long lived α ray emitters

The systematic investigations by H Geiger and J M Nuttall are summarized in their classical form as the Geiger Nuttall diagram, Fig 6 Here the half period T is represented by the radioactive de cay constant $\lambda = 0.693/T$, and the α ray energy E is represented by the range R of the α rays in centi meters of air, Geiger having shown previously that the range of an α ray is closely proportional to the three halves power of its kinetic energy, that is

$$R = \text{const} \times E^{3/2} \text{ (Geiger's rule)} \qquad (17)$$

The members of the three naturally occurring de cay series are seen to follow a clear relationship of decreasing half period (increasing λ) with increas ing energy (range)

The Geiger Nuttall rule is inexplicable by clas sical physics but emerges clearly from quantum or wave mechanics In 1928 the hypothesis of

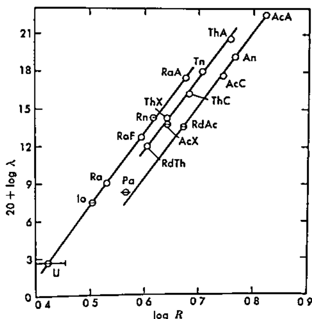


Fig 6 Original form of the Geiger Nuttall relation ship between α ray disintegration energy, expressed as range R of the α rays, and half period expressed as decay constant λ (From H Geiger, Z Physik 8 45–57, 1921)

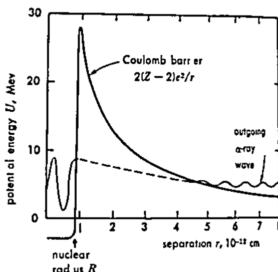


Fig 7 Schematic of nuclear potential barrier, illus trating emission of an α ray as a wave which can be transmitted through the barrier

transmission through nuclear potential barriers, as introduced by G Gamow and independently by R W Gurney and E. U Condon, was shown to give a satisfactory account of the α-decay data, and it has been altered subsequently only in details. The form of the barrier penetration equations is such that correlation plots of $\log \lambda$ against $1/\sqrt{E}$ give nearly straight lines

Nuclear potential barrier. At distances r which are large compared with the nuclear radius, the po tential energy of an α ray, whose charge is $2e$, in the field of a residual nucleus, whose charge is $(Z-2)e$, is $2(Z-2)e^2/r$ At very close dis tances, this electrostatic repulsion is opposed and overcome by short range specifically nuclear at tractive forces The net potential energy U, as a function of the separation r between the α ray and its residual nucleus is called the nuclear potential barrier

One of several operating definitions of the nuclear radius R is the distance $r = R$ at which the attractive nuclear forces just balance the repulsive electrostatic forces At this distance, called the top of the nuclear barrier the potential energy is about 25–30 Mev for typical cases of heavy, α emitting nuclei as indicated in Fig 7 *See* POTENTIAL BAR RIER *see also* NUCLEAR STRUCTURE, QUANTUM ME CHANICS

Inside the nucleus the α particle is represented as a de Broglie matter wave According to wave mechanics this wave has a very small but finite probability of being transmitted through the nu clear potential energy barrier and thus of emerging as an α ray emitted from the nucleus The trans-

ergy barrier is analogous to the familiar case of the transmission of ordinary visible light through an opaque metal such as gold, if the gold is thin

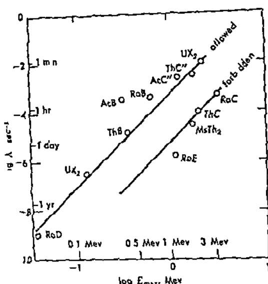


Fig 12 Sargent diagram for some naturally occurring β ray emitters Maximum energy of β ray spectrum E_{max} is plotted against partial decay constant λ for principal branch of β-decay half period T is also shown (From B W Sargent Proc Roy Soc (London) A139(839) 659–673 1933)

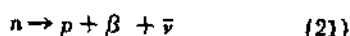
light (3×10^{10} cm/sec) and is emitted as a companion particle with each β ray

Two forms of neutrino are distinguished In positron β-decay a proton p in the nucleus transforms into a neutron n in the nucleus thus reducing the nuclear charge by one unit At the time of this transition two particles the positron β^+ and the neutrino ν are created and emitted The emitted β and ν together carry away the energy E_{max} of the transition and provide for conservation of energy momentum angular momentum charge and statistics Thus positron β decay is represented by



$$p \rightarrow n + \beta^+ + \nu \qquad (20)$$

Negatron β decay is a closely related process except that a neutron n changes to a proton p in the nucleus and a negatron β ray β and its characteristic companion particle the antineutrino $\bar{\nu}$ are emitted Thus



$$n \rightarrow p + \beta^- + \bar{\nu} \qquad (21)$$

The antineutrino is the antiparticle of the neutrino They have the same properties of zero charge and zero rest mass and differ only with respect to the direction of alignment of their intrinsic spin along their direction of motion In most contexts the term neutrino includes both its forms neutrino and antineutrino

The interaction of neutrinos with matter is exceedingly feeble A neutrino can pass all the way *through* the sun with little chance of a collision. The *thickness* of lead required to attenuate neutrinos by the factor ½ is about 10^{20} cm or 100 light years of lead[1] See NEUTRINO see also ANTIMATTER POSITRON

Average beta energy. Charged particles, such as β rays or α rays are easily absorbed in matter and their kinetic energy is thereby converted into heat In β decay, the average energy E_{av} of the β rays is far less than the maximum energy E_{max} of the particular β ray spectrum For example, *in* the case of RaE integration over the energy spectrum ... and hence E_{av} ... 28% of the ... is available as kinetic energy of β rays This is in agreement with direct measurements on RaE The remaining 72% of the disintegration energy of RaE is emitted as kinetic energy of neutrinos and is not recoverable in finite absorbers

The detailed shape of β ray spectra and hence the exact value of the ratio E_{av}/E_{max} varies somewhat with Z, E_{max} the degree of forbiddenness of the transition and the sign of charge of the emitted β ray A rough rule of thumb which covers many practical cases is $E_{av} = (0.40 \pm 0.05)E_{max}$.

Fermi theory By postulating the simultaneous emission of a β ray and a neutrino, as in Eq (20), E Fermi developed in 1934 a quantum mechanical theory which satisfactorily gives the shape of β ray spectra (as in Figs 10 and 11) and the relative half periods of β ray emitters (as in Fig 12)

The energy distribution of β rays in allowed transitions is then given by

$$N(W)\,dW = \frac{|P|^2}{\tau_0} F(Z,W)(W^2-1)^{1/2}(W_0-W)^2 W\,dW \qquad (22)$$

where $N(W)\,dW$ = number of β rays in energy range W to $W + dW$

$W = 1 + E/(m_0c^2)$ = total energy of β ray in units of rest energy m_0c^2 = 0.51 Mev for an electron (m_0 = electron mass c = velocity of light)

$W_0 = 1 + E_{max}/(m_0c^2)$ = maximum energy of the β ray spectrum

τ_0 = time constant $\cong$ 7000 sec

$F(Z,W)$ = complex, dimensionless function involving the nuclear radius nuclear charge and β ray energy $F(Z,W)$ has been evaluated and published in tabular form by the National Bureau of Standards as *Tables for the Analysis of Beta Spectra* Appl Math Ser 13 1952

Physically this distribution function involves the product of the energy W and momentum $(W^2-1)^{1/2}$ *of the* β ray times the energy $(W_0 - W)$ and the momentum $(W_0 - W)/c$ of the neutrino

individual points show the measured half periods and α disintegration energies (α ray energy plus recoil energy) for a number of heavy emitters of α rays The smooth curves are drawn using the wave mechanical theory of transmission through nuclear barriers, with a nuclear radius of $R = 1\ 48 \times 10^{-13}A^{1/3}$ cm, where A is the mass number of the α ray decay product The agreement between experiment and theory is good

BETA-RAY DECAY

Beta ray decay is a type of radioactivity in which the parent nucleus emits a β ray There are two types of β decay in negatron β decay (β^-) the emitted β ray is a negatively charged electron (negatron), in positron β decay (β^+) the emitted β ray is a positively charged electron (positron) In β decay the atomic number shifts by one unit of charge, while the mass number remains unchanged (Table 1)

The number vs energy distribution of emitted β rays is continuous, as illustrated in Figs 9–11 For each β ray emitter there is a definite maximum

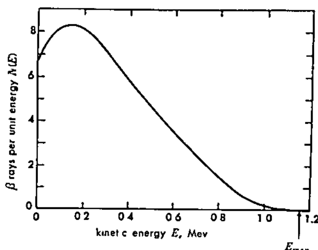


Fig 9 Energy spectrum of negatron β rays from RaE (*From G J Neary, Proc Roy Soc (London), A175(960) 71–87, 1940*)

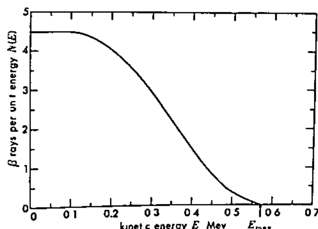


Fig 10 Energy spectrum of negatron β rays from Cu^{64} (*From R D Evans, The Atomic Nucleus, McGraw Hill 1955*)

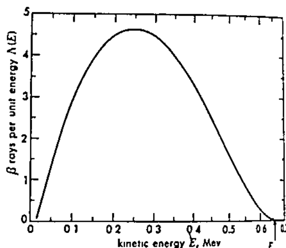


Fig 11 Energy spectrum of positron β-rays from Cu^{64} (*From R D Evans, The Atomic Nucleus McGraw Hill, 1955*)

or upper limit to the energy spectrum of β rays This maximum energy, E_{max}, corresponds to the change in nuclear energy in the β decay Thus $E_{max} = 1\ 16$ for RaE, $E_{max} = 0\ 57$ Mev for β^- decay of Cu^{64} and $E_{max} = 0\ 66$ Mev for β^+ decay of Cu^{64} As in the case of α decay some β ray spectra are not this simple but include additional continuous spectra having less maximum energy and leaving the product nucleus in an excited level from which γ rays are then emitted

Sargent diagrams. For a given class of β emitters the half period decreases rapidly as the decay energy E_{max} increases This systematic behavior shown in Fig 12, was first pointed out by B W Sargent and is the β decay analog of the Geiger Nuttall rule (Fig 6) for α decay

The division of β ray transitions into many classes as allowed first forbidden, second forbidden third forbidden etc was first made on empirical grounds For the same E_{max}, the half period increases by roughly a factor of 100 for each increase of one degree of forbiddenness Figure 12 shows allowed and (first) forbidden transitions Theory now shows that the allowed transitions are those in ... nuclei

or differs by only one unit of angular momentum. Transitions of various degrees of forbiddenness result from various combinations of the parity and angular momentum of the parent and product nuclei as described by the so called selection rules for β decay *See* PARITY (QUANTUM MECHANICS) SELECTION RULES (PHYSICS)

Neutrinos The continuous spectrum of β ray energies shown in Figs 9–11 implies the simultaneous emission of a second particle besides the

and the β ray always equals E_{max} for the particular transition involved The neutrino has zero rest mass zero charge travels at the same speed as

lati
ent
for

cleus ...

and a classification of γ ray transitions as "electric" or as "magnetic" multipole radiation of multipole order 2^l, where $l = 1$ is called dipole radiation, $l = 2$ is quadrupole radiation, and $l = 3$ is octupole, l being the vector change in nuclear angular momentum. The most common type of γ ray transition in nuclei is the electric quadrupole ($l = 2$).

Mean life for transitions. A reasonably successful approximate theory of the mean life for γ ray decay was developed by V. F. Weisskopf in 1951 using the single-particle shell model of nuclei. Figure 14 summarizes the numerical consequences of this theory. Note that an electric quadrupole ($l = 2$) transition of about 1 Mev is expected to take place with a mean life, τ_{el}, or mean delay in the upper level of about 10^{-11} sec. Thus most γ ray transitions are prompt transitions, in which the mean life of the excited level is too short to be measured easily. Figure 14 is for electric multipole transitions. The mean life τ_{mag} for magnetic multipoles is of the order of 30 (for $A = 20$) to 150 (for $A = 200$) times longer than τ_{el}. *See* MULTIPOLE RADIATION.

Internal conversion. An alternative type of de-excitation which always competes with γ ray emission is known as *internal conversion*. Instead of the

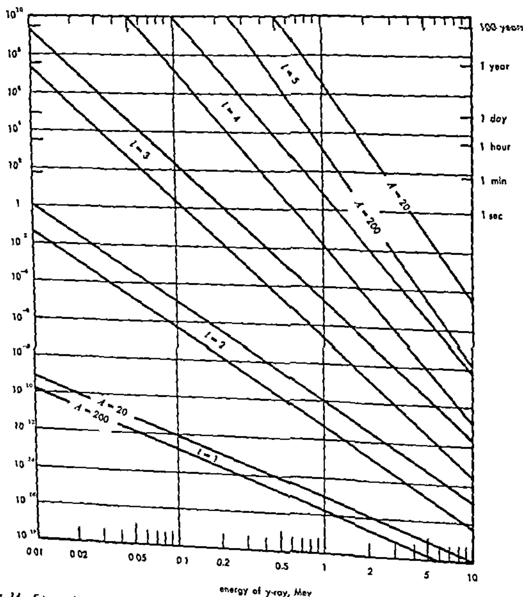


Fig. 14. Estimated mean life τ_{el} of nuclear levels for deexcitation by emission of electric multipole γ-rays of order 2^l. Influence of atomic weight or mass number is indicated by curves for mass $A = 200$ and $A = 20$ plotted for each value of l. (From R. D. Evans, *The Atomic Nucleus*, McGraw-Hill, 1955)

The half period T of β decay can be derived from Eq (22) because the radioactive decay constant $\lambda = 0.693/T$ is simply the total probability of decay, or $N(W)\,dW$ integrated over all possible values of the β ray energy from $W = 1$ to $W = W_0$

Equation (22) matches the energy spectra of allowed β ray transitions and therefore furnishes one type of experimental verification of the properties of neutrinos Its counterpart in terms of the β ray momentum spectrum is often used for the analysis of spectra and is

$$N(\eta)\,d\eta = \frac{|P|^2}{\tau_0} F(Z,\eta)(W_0 - W)^2\eta^2\,d\eta \qquad (23)$$

where $N(\eta)\,d\eta$ = number of β rays in the momentum interval from η to $\eta + d\eta$
$\eta = (W^2 - 1)^{1/2}$ = *momentum of the* β ray in units of m_0c
$F(Z,\eta) = F(Z,W)$ of Eq (22)

The momentum distribution is much more nearly symmetric than its corresponding energy spectrum

Kurie plots For allowed transitions the transition matrix element $|P|$ is independent of the momentum η Then Eq (23) can be put in the form

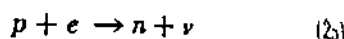
$$\left[\frac{N(\eta)}{\eta^2F(Z,\eta)}\right]^{1/2} = \text{const}\,(W_0 - W) \qquad (24)$$

Therefore a straight line results when the quantity $\sqrt{N/\eta^2F}$ is plotted against β ray energy either as W or as E on a linear scale Such graphs are called Kurie plots, or Fermi plots They are especially useful for revealing deviations from the theory and for obtaining the upper energy limit E_{max} as the extrapolated intercept of $\sqrt{N/\eta F}$ on the energy axis Practically all new results on the shape of β ray spectra are published as Kurie plots rather than as actual momentum or energy spectra

Figure 13 shows a representative Kurie plot for In^{114} Note that N means $N(\eta)$ the number of β rays in a momentum interval of constant spread $\Delta\eta$ The horizontal scale is the kinetic energy E which corresponds to the midpoint of the momentum interval η to $\eta + \Delta\eta$ When spectral data give a straight line such as this one then $N(\eta)$ is in agreement with the Fermi momentum distribution Eq (23), and the intercept of this straight line, on the energy axis gives the disintegration energy E_{max}

Electron capture transitions Whenever it is energetically allowed by the mass difference between neighboring isobars a nucleus Z may capture one of its own atomic electrons and transform to the isobar of atomic number $Z - 1$ (Table 1) Usually the electron capture (EC) transition involves an electron from the K shell of atomic electrons, because these innermost *electrons have the greatest* probability density of being in or near the nucleus *See* ELECTRON CAPTURE

In EC transitions a proton p bound in the parent nucleus absorbs an electron e and changes to a bound neutron n The disintegration energy is carried away by an emitted neutrino ν



$$p + e \rightarrow n + \nu \qquad (25)$$

The residual nucleus may be left either in its ground level or in an excited level from which γ ray emission follows

EC transitions compete with all cases of positron β ray decay EC has an energetic advantage over β^+ decay equivalent to the mass of two electrons, or 1.02 Mev, because in Eq (25) one electron mass e enters the reaction and is available whereas in Eq (20) one electron mass β must be produced as a product of the positron β ray decay For example, in the radioactive decay of $_{29}Cu^{64}$ twice as many transitions go by EC to $_{28}Ni^{64}$ as go by positron β decay to the same decay product. In the heavy, high Z elements, EC is greatly favored over the competing β^+ decay and examples of measurable β^+ decay are practically unknown for Z greater than 80 although there are a large number of examples of electron capture

Several examples are known of completely pure EC radioactivity in which there is insufficient nuclear energy to allow any positron β ray decay For example, $_{26}Fe^{55}$ emits no positron β rays but transforms with a half period of 2.6 years entirely by EC to the ground level of $_{25}Mn^{55}$ This radioactivity is detectable through the K series x rays which are emitted from Mn^{55} when the atomic electron vacancy, produced by nuclear capture of a K electron, refills from the L shell of atomic electrons *See* X RAY(S), PHYSICAL NATURE OF

GAMMA-RAY DECAY

Gamma ray decay is a transition between two excited levels of a nucleus or between an excited level and the ground level A nucleus in its ground *level cannot emit any γ radiation Therefore* γ ray decay occurs only as a sequel of those instances of α decay β decay or electron capture in which the product nucleus is left in an excited energy level

A γ ray is an electromagnetic radiation (photon) in the same family with radio waves, visible light and x rays (*see* ELECTROMAGNETIC RADIATION) The energy of a γ ray is $h\nu$ where h is Planck's constant and ν is the frequency of oscil

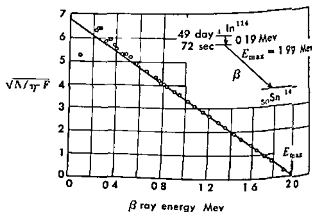


Fig 13 Kurie plot of negatron β ray spectrum of In^{114} (From J L. Lawson and J M Cork Phys Rev 57(11) 982–994, 1940)

und level of Ba^{137} by $l = 4$ units of angular mentum has a half period of 2 6 *min* and de s to the *ground* level of Ba^{137} by delayed 6-Mev γ rays and a spectrum of conversion ctrons

Most of the isomers occur in nuclei which have d mass number *A* Then either the number of otons *Z* in the nucleus is odd or the number of utrons *N* in the nucleus is odd The frequency stribution of odd *A* isomeric pairs excited level d ground level displays *so-called islands of imerism in which the odd proton* or odd neutron mber is less than 50 or less than 82 as shown Fig 15 This distribution is one of several es of evidence for closed shells *of identical nu eons* at *N* or *Z* 50 or 82 in nuclei and it ays an important role in the so-called shell odel of nuclei *See* ISOMERISM NUCLEAR [R D E]

Bibliography J M Blatt and V F Weisskopf *heoretical Nuclear Physics* 1952 R D Evans ... *Intro*

Radioactivity (applications)

The applications of radioactivity include the use of radioactive isotopes as tracers to study chemical and physical substances *and as modifiers to change* the characteristics of substances

Most artificially produced radioactive isotopes (radioisotopes) emit electrons and some emit γ rays (*see* RADIOACTIVITY) This *radiation is* identical to that produced by many high energy particle accelerators and can be used to produce radiation changes in materials Radioactive iso

isotope and its radiation tracked as it mixes within it These radioactive tracers have many biological medical and industrial uses *See* TRACER RADIOACTIVE *see also* RADIOISOTOPE RADIOISOTOPE (BIOLOGY) RADIOISOTOPE PRODUCTION

Radioactivity in biology If given in sufficient intensity radiation will produce marked biological changes Most of these will be *detrimental* to the organism but some may actually produce beneficial effects

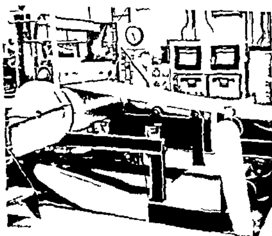
Radiation is widely used to induce mutation in living organisms New varieties of grains and other plants have been produced in this manner A particularly important application has been the irradiation of yeast to produce new forms of penicillin Radiation has also been used to control the population of *insect pests by introducing* radiation sterilized groups into the population *See* RADIATION BIOLOGY

Radioactivity in medicine The use of radioiso-

...ortion of the function of the thyroid gland in cases where the thyroid gland is overactive It is also used in treating certain types of thyroid carcinoma Gold 198 *injected in colloidal form directly into tumor* tissue has also achieved considerable success in tumor therapy (*see* IRRADIATION ISOTOPIC) The intravenous injection of phosphorus-32 is a standard method of treating leukemia and other blood diseases *Successful* as it has been *in the past* medical therapy with radio *active* isotopes still leaves much to be desired The ideal isotope therapy procedure would use an isotope which is highly specific for the tissue to be destroyed and which avoids *all other tissues* Present techniques are far from approximating this ideal However as increased knowledge of the growth and structure of tumor tissue becomes available it is certain that new and more powerful isotope therapy techniques will evolve *See* ISOTOPE DILUTION TECHNIQUES

Radioactive isotopes are also used as an alternative to high voltage machines as *radiation sources* Cobalt 60 the most widely used isotope for this purpose *is encapsulated* and surrounded by a heavy lead shield An opening in this shield permits an intense beam of radiation to emerge This beam can be used for external radiation therapy

Radioactivity in industry *Radioisotopes as* radiation sources now perform many tasks in industrial research and development process control and monitoring One such application is the measurement of thickness of paper metal or other materials *in a continuous production process In a* typical application a source of β emitting isotopes is mounted beneath the strip of material for which the thickness is desired The detector usually an ionization chamber is mounted above the material and records the amount of *radiation* which penetrates the strip (see illustration) The intensity transmitted through the strip is related to the thickness of the material Thickness gages of this type may be used *to record the thickness of the*



Arrangement for automatic control of thickness of plastic sheet Two thickness gages measure thickness continuously and control apparatus regulating production of plastic sheet (Tracerlab Inc)

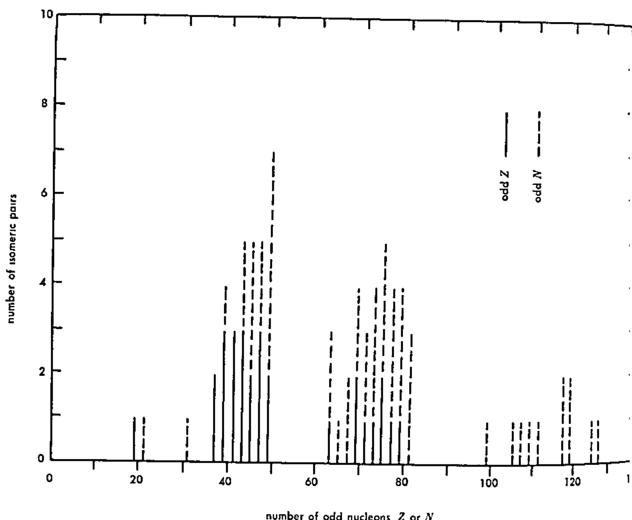


Fig 15 Frequency distribution of odd A isomeric pairs displays islands of isomerism at neutron numbers N and proton numbers Z which are less than 50, or less than 82 (*From R D Evans, The Atomic Nucleus McGraw Hill, 1955*)

emission of a γ ray the nuclear excitation energy can be transferred directly to a bound electron of the same atom Then the nuclear energy difference is converted to energy of an atomic electron, which is ejected from the atom with a kinetic energy E_i given by

$$E_i = W - B_i \qquad (26)$$

where B_i is the original atomic binding energy of the particular electron, which is ejected and W is the nuclear transition energy which would otherwise have been emitted as a γ ray photon having energy $h\nu = W$

The spectrum of internal conversion electrons is then a series of discrete energies or "lines" corresponding to the individual values of B_i for the various $K, L, M,$, electrons in the atom From the spacing of the E_i values in this conversion electron spectrum it is possible to assign definitely the atomic number Z of the atom in which the nuclear transition W took place In this way it is known that the competing γ ray emission is a sequel and not an antecedent of α decay, β decay, and electron capture transitions

The internal conversion coefficient α is the ratio of the number of transitions proceeding by internal conversion to the number going by γ ray emission, for any particular nuclear transformation from an excited level to a lower lying level In general, this probability of internal conversion relative to γ ray emission increases with increasing atomic number Z with increasing multipole order 2^l and with decreasing nuclear deexcitation energy W In middle weight elements for $W = 1$ Mev α is of the order of 10^{-3}–10^{-4} while for $W = 0.$ Mev, α is of the order of 0 1 for electric $l =$ transitions and 10 or larger for electric l transitions

Isomeric transitions. Measurably delayed radioactive transitions from an excited level of a nucleus rather than from the ground level are known as isomeric transitions The measurably long lived excited level is called an isomeric level or an isomer of the ground level

Figure 14 shows that if the excitation energy is small (say 0 5 Mev or less) and the angular momentum difference l is large (say $l = 3$ or more) then the mean life of an excited level for γ ray or conversion electron emission can be of the order of 1 sec up to several yr

A typical isomeric transition occurs in $_{56}Ba^{137}$ where an excited level at 0 66 Mev is produced in about 92% of the β ray transitions of $_{55}Cs^{137}$ to $_{56}Ba^{137}$ This 0 66 Mev level differs from the

Thus standards of radium in which the half life is about 1620 years, have an almost permanent value Standards of radioisotopes with very much shorter half lives of the order of a day require preparation and calibration immediately prior to their use. See CURIE [L F CS]

Bibliography W B Mann The preparation and maintenance of standards of radioactivity *Intern J Appl Radiation and Isotopes* 1 3–23, 1956

Radiocarbon dating

A method of determining the absolute age of car bon bearing materials which have formed in equilib- ... 70 000

decay of the cosmic ray produced isotope ca bon 14 (C^{14}) It has been successfully applied to such materials as wood plant remains charcoal ... shell and fresh water carbonate deposits

logic events and of the development ... man

RADIOACTIVE DECAY OF CARBON-14

giving the atom the chemical characteristics of nitrogen As with all radioactive isotopes C^{14} may be characterized by a half life (5570 ± 40 years) the time required for half of a given number of atoms to undergo radioactive decay Since the rate of transformation cannot be altered by any physical conditions on the surface of the earth the rate of disappearance of C^{14} from a sample bears an absolute relationship to time

Since the earth is about 4.5×10^9 years old any C^{14} produced initially with the rest of the isotopes making up the solar system would have long since disappeared See EARTH (AGE OF) C^{14} is observed on the surface of the earth only because it is continuously being produced Primary cosmic ray particles upon entering the atmosphere produce neutrons which react with the abundant N^{14} atoms of the air A neutron enters the N^{14} nucleus and knocks out a proton converting the atom to C^{14} which rapidly is oxidized to a $C^{14}O_2$ molecule If this production has proceeded at a constant rate for many thousands of years then the amount of C^{14} present on the surface of the earth should reach a constant value This can be most easily understood by considering the flow of water through a funnel If water is poured into a funnel at a uniform rate the level or amount of water in the funnel will increase to some constant amount and remain unchanged as long as the rate of pouring is not altered Likewise, since the rate of decay of C^{14} atoms is a function only of the number of atoms present the total number of C^{14} atoms in the reservoir (atmosphere, biosphere and hydrosphere) must be constant if the production rate does not change See RADIOACTIVE SPECIES PRODUCED BY COSMIC RAYS

The C^{14} produced in the atmosphere gradually ...

of C^{14} in the carbon in the CO_2 of the ... atmosphere in the earth's hydrosphere (largely as bicarbonate ion), and in the earth's biosphere as organic carbon is nearly uniform When a material which received its carbon from this dynamic reservoir is isolated its C^{14} concentration begins to decrease at a rate of half every 5570 years Thus, if the C^{14} concentration in the carbon from a plant of unknown age were found to be half that in a plant living today, its age would be estimated as 5570 years

MEASUREMENT OF RADIOCARBON

The abundance of radiocarbon in contemporary materials is so small that direct detection is impossible. Only one carbon atom in 10^{12} in living wood for example is C^{14} This is far below the sensitivity limit of an isotope measurement device such as the mass spectrograph Fortunately the presence of C^{14} can be demonstrated through its radioactivity The beta particle which is emitted during the nuclear transformation process can be detected with a

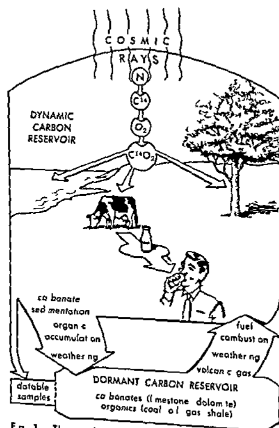


Fig 1 The earth's carbon

material continuously, or they may be used as a component of a feedback system to maintain the thickness at a desired value Different isotope sources are used, depending on the range of thick nesses which must be measured A soft (low en ergy) β emitter would be used for thickness gaging of paper, while a hard or high energy β emitter might be used for certain metal sheets

Another important industrial application of ra dioactivity is in the testing of metal castings for flaws In this procedure, photographic film is at tached to the casting and a radioisotope source, usually cobalt 60, is placed nearby Flaws within the casting are shown as lines on the photographic film This type of radiographic testing has become a standard procedure in the production of metallic castings which must withstand extremes of tem perature and pressure

The possibility of using radioactivity to modify the properties of manufactured goods offers con siderable promise, although few applications have yet reached the production stage Polyethylene, for example, when irradiated in an intense γ ray flux, exhibits marked improvement in its stability to heat This modification in the polyethylene is caused by radiation induced cross linkages within the polymer structure By and large, however, the effect of high levels of radiation on most materials is detrimental rather than beneficial

Radioactivity also offers considerable promise in the food industry for sterilization of various food products The sterilization of meat and other per ishable commodities has received considerable study In the drug industry, certain biologicals are currently being sterilized by ionizing radia tion

Radioactivity in analytical chemistry. Radioac tive isotopes can be used to determine rates of reactions and reaction kinetics They are particu larly valuable in analytical chemistry in the deter mination of very small quantities of certain ele- ments by the process of activation analysis In this technique, the unknown material is exposed to a beam of particles, for example the material can be placed in an intense thermal neutron flux within a reactor The interaction of neutrons with various nuclei in the sample produces radioactive iso- topes Following the removal of the sample these isotopes can be measured by β or γ ray detecting devices If the element to be measured produces radioactive isotopes with a sufficient radiation yield extremely small (millimicrogram) quanti ties can be assayed

In most cases, the measurement is standardized by irradiating known concentrations of the desired material and measuring this standard under identi cal conditions as with the unknown The actual limit of sensitivity in this technique is often set by trace amounts of the chemical to be assayed pres- ent in the container or in other chemicals used in preparing the sample For additional information, see ACTIVATION ANALYSIS *See also* AGRICULTURAL SCIENCE (ANIMAL), GEOCHRONOMETRY, LUMINOUS

Elements which may be determined in trace quantities by neutron activation analysis*

Elements	Sensitivity of detection,† μg
Eu	10^{-6}
Mn, In Dy, Ho, Re, Ir, Au	10^{-5}
Na, Cu As Pd I, La Pr Sm Yb Lu, Ta W	10^{-4}
Cl Sc, Ga, Ge, Se Br, Rb Y Sb, Ba, Gd, Tb, Er, Tm, Os, U	10^{-3}
P, Cr, Co, Ni Zn, Sr, Ru, Cd Sn, Te, Ce, Nd, Hf, Pt, Hg, Th	10^{-1}
Si, K, Zr, Mo Ag Cs, Tl Bi	1
S, Ca Fe, Pb	1 100

* Taken from G H Morrison, Neutron activation anal ysis for trace elements, *Appl Spectroscopy* 10(2) 1956.

† The amount of an element that would produce suffi cient activity to be measured after the element has been exposed to a flux of 3.4×10^{12} neutrons/(cm²)(sec) for 3 days The sensitivities for elements whose radioactive iso- topes have very short half lives will decrease considerably with the length of time between removal from the pile and measurement of activity Isotopes with half lives less than 10 min have been omitted

PAINT, PHOTOSYNTHESIS, PLANT, MINERAL NUTRI TION OF, PROSPECTING, RADIOCHEMISTRY, WELL LOGGING (MINERAL) [C.L.L.]

Bibliography C. L. Comar, *Radioisotopes in Biology and Agriculture, Principles and Practice*, 1955, *The Effects of Atomic Radiation on Oceanog raphy and Fisheries*, Natl Acad Sci Natl Re search Council Publ 551, 1957, W E Siri *Iso- topic Tracers and Nuclear Radiations* 1949 United Nations *Proc Intern Conf Peaceful Uses Atomic Energy*, vols 10–15, 1956

Radioactivity standards

Calibrated standard sources of radioactive sub-

in terms of some figure proportional to this num- ber The calibration of the standard source in terms of the number of radioactive atoms is usually an elaborate procedure but need only be carried out once and the calibration may be made at a stand

tor of radiation to the radiation from a sample and from the standard In each case the intensity of the

cause this number is also proportional to the dis- integration rate of a source *See* HALF LIFE, RADIO- ACTIVITY

The number of situations in which standards of radioactivity can be used is limited by a number of factors For example the radioactive species in- volved must have a half life long enough that the standards will continue to have sufficient activity for the period of time in which they are to b ed

record the mesons entering the su... electronic cancellation of any pulses produced by mesons in the sample counter (see Fig 2) A fur ther reduction in the background is possible in some cases through the use of an internal antico incidence system See LOW LEVEL COUNTING

The characteristics of a typical high sensitivity s counter are compared with those of the original F Libby black carbon counter in Table 1 Such ctors as sample size precision on contemporary aterials and ultimate sensitivity for age dating e also summarized It should be pointed out that hereas the limitation of counting sensitivity al ws samples up to only 50 000 years in age to be ... the

nding the potential age range to 70 000 years

Table 1 Characteristics of C^{14} counters

	Black carbon counter	Gas proportional counter
Sample size moles carbon	0 5	0 3
Contemporary sample counts per min	5 5	40
Unshielded background counts per min	200	600
Shielded background counts per min	4	3
Precision for contemporary samples, %	±3	±0 5
Age limit, years	25 000	50 000

LIMITATIONS OF THE RADIOCARBON METHOD

The sources of error in a radiocarbon age deter mination can be divided into two groups errors in the measurement and errors in the assumptions

Errors in the measurement The measurement error results largely from the statistical error in the counting Just as 10 flips of a coin will not al ways produce 5 heads and 5 tails the same number of radioactive transformations will not always oc cur in a given sample in one 5-min period as in the next As in the case of the coin the deviations are random and occur with a probability which can be mathematically predicted Also in both cases the deviations will decrease as the number of events (hence radioactive transformations) observed in creases. Thus most radiocarbon activity measure ments are carried out for periods of from 1–4 days Reduction of the error by counting for longer pe riods is impractical

The counting error is usually reexpressed as an age error (that is, 11 000 ± 200 years) The error represents the one-sigma level which means that there is a 67% probability that the C^{14} age lies within the stated limits. There is a 95% probability

Table 2 Interlaboratory comparisons

Laboratory	Sample number	Age*
Groningen	GRO 1172	2885 ± 60
Groningen	GRO 1512	2770 ± 50
Columbia University	L 427	2700 ± 90
University of Michigan	M 290	5130 ± 150
Columbia University	L 214	5090 ± 300
Copenhagen	K 101	10 890 ± 210
U S Geological Survey	W 82	10 260 ± 200
U S Geological Survey	W 84	10 510 ± 180
Heidelberg	H 105–87	11 500 ± 300
Columbia University	L 289M	11 700 ± 200
U S Geological Survey	W 442	12 050 ± 400
Columbia University	L 183B	28 200 ± 1000
U S Geological Survey	W 83	27 500 ± 1200

* Years before present

that the C^{14} age does not lie outside the two-sigma level (twice the stated error) When the radioactiv ity of the sample is less than twice the counting error a minimum age is given (that is $\geqq$37,000 years)

Other sources of laboratory error caused by var iation in counter efficiency, variations in counter background and contamination by extraneous

...

black carbon measurements Remeasurements by the more precise gas counting techniques have shown that some of these measurements were seri ously in error Air borne fission products easily picked up by the highly absorbent dry carbon dur ing preparation were shown to be responsible for many of these errors Many interlaboratory com parisons such as those shown in Table 2 indicate that such errors are quite infrequent when the gas counting techniques are used The statistical errors as a function of age for a typical gas counter are shown in Fig 3

Errors in the assumptions Deviations from the basic assumptions of the method limit the absolute accuracy which can be attained These assumptions are as follows (1) the initial C^{14} activity in the

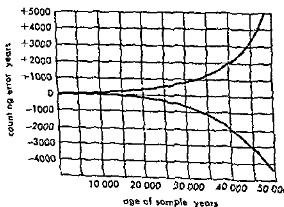


Fig 3 Measurement error for sensitive gas counting system plotted as function of sample age

Geiger counter Since the number of beta particles emitted by a carbon sample in a given interval of time is directly proportional to the number of radiocarbon atoms present, an age determination can be made by comparing the radioactivity of carbon from a sample of unknown age with that from a contemporary sample The relationship between the age t and the radioactivities is

$$t = 8040 \ln (A_0/A) \quad \text{years}$$

where A_0 and A are respectively, the radioactivity of the carbon prepared from the contemporary and from the unknown sample

In order to obtain precise measurements of the radioactivity of carbon very elaborate chemical and radiochemical procedures had to be developed This is because the number of beta particles to be detected is extremely small (13 disintegrations per minute per gram contemporary carbon) and because they lack penetrating power (maximum energy 0 15 Mev) In order to record these radiations the sample has to be placed within an extremely sensitive radiation counter which has been adequately shielded from extraneous radioactivity (Fig 2)

The original measurements were made by converting the carbon in the sample to carbon black The sample was mounted on the inside of a steel cylinder which could be inserted into a screen wire Geiger counter Subsequently it was found that more precise measurements could be made if the sample was converted into a gas such as carbon dioxide, acetylene, or methane The sample in this case is used as the filling gas for the counter

The reduction of background radiation resulting from cosmic rays and natural radioactivity in the materials which make up the laboratory and its surroundings is accomplished as follows. The counter itself is constructed of materials as free as possible of radioactivity The bulk of the external gamma radiation is eliminated by placing the counter within an iron shield from 8 to as much as 30 in. thick Additional mercury shielding is often placed immediately around the counter to eliminate gamma radiation originating in the iron shield itself H DeVries has shown that the addition of

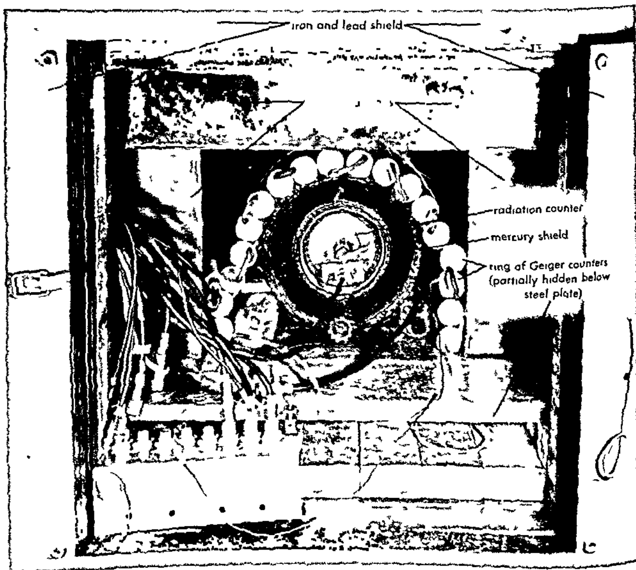


Fig 2 Large volume proportional counter used for carbon 14 measurements Outer shield is closed by rolling doors Sample is introduced into radiation counter in form of carbon dioxide gas (Geochemical Laboratory Lamont Geological Observatory, Columbia University)

record the *mesons entering* the shield ... electronic cancellation of any pulses produced by mesons in the sample counter (see Fig 2) A further reduction in the background is possible in some cases through the use of an internal anticoincidence system *See* LOW LEVEL COUNTING

The characteristics of a typical high sensitivity gas counter are compared with those of the original W F Libby black carbon counter *in Table 1 Such* factors as sample size, precision on contemporary materials, and ultimate sensitivity for age dating are also summarized It should be pointed out that whereas the limitation of counting sensitivity allows samples up to only 50 000 years in age to be [illegible] directly preenrichment of the C^{14} in the [illegible]le gas by thermal diffusion raises the experi[illegible]al sensitivity by as much as a factor of 16 ex[illegible]ing the potential age range to 70 000 years

[illegible] 1 Characteristics of C^{14} counters

	Black carbon counter	Gas proportional counter
[illegible]le size, moles carbon	0 5	0 3
[illegible]emporary sample, [illegible]unts per min	5 5	40
[illegible]ielded background [illegible]unts per min	200	600
[illegible]lded background, [illegible]unts per min	4	3
[illegible]sion for contemporary [illegible]mples, %	±3	±0 5
[illegible]limit, years	25 000	50 000

LIMITATIONS OF THE RADIOCARBON METHOD

The sources of error in a radiocarbon age determination can be divided into two groups errors in [illegible] measurement and errors in the assumptions

Errors in the measurement. The measurement [illegible]or results largely from the statistical error in [illegible]e counting Just as 10 flips of a coin will not al[illegible]ys *produce* 5 heads and 5 tails the same number [illegible] radioactive transformations will *not* always oc[illegible]r in a given sample in one 5 min period as in the [illegible]st. As in the case of the coin the deviations are [illegible]ndom and *occur* with a probability which can be [illegible]athematically predicted. Also *in both* cases the [illegible]eviations will decrease as the number of events [illegible]hence radioactive transformations) observed in[illegible]reases. Thus *most* radiocarbon activity measure[illegible]ments are carried out for periods *of from* 1–4 days [illegible]eduction of the error by counting for longer pe[illegible]ds is impractical

The counting *error* is usually reexpressed as an [illegible]ge error (that is, 11 000 ± 200 years) The error

Table 2. Interlaboratory comparisons

Laboratory	Sample number	Age*
Groningen	GRO 1172	2885 ± 60
Groningen	GRO 1512	2770 ± 50
Columbia University	L 427	2770 ± 90
University of Michigan	M 290	5130 ± 150
Columbia University	L 214	5090 ± 300
Copenhagen	K 101	10 890 ± 210
U S Geological Survey	W 82	10 260 ± 200
U S Geological Survey	W 81	10 510 ± 180
Heidelberg	H 105–87	11 500 ± 300
Columbia University	L 289M	11 700 ± 200
U S Geological Survey	W 412	12 050 ± 100
Columbia University	L 185B	28 200 ± 1000
U S Geological Survey	W 83	27,500 ± 1200

* Years before present

that the C^{14} *age* does not lie outside the two-sigma level (twice the stated error) When the radioactivity of the sample is less than twice the counting error a minimum age is given (that is ≧37,000 years)

Other sources of laboratory error caused by variation in counter efficiency, variations in counter background and contamination by extraneous radioactivity are generally small with respect to *the counting error but may cause significant* deviations in some cases This is especially true for the black carbon measurements. Remeasurements by the more precise gas counting techniques have shown that some of these measurements were seriously in error Air borne fission products easily picked up by the highly absorbent dry carbon during preparation were shown to be responsible for many of these errors Many interlaboratory comparisons *such as those shown in* Table 2 *indicate* that such errors are quite infrequent when the gas counting techniques are used The statistical errors as a function of age for a typical gas counter are *shown in Fig 3*

Errors in the assumptions. Deviations from the basic assumptions of the method limit the absolute accuracy which can be attained These assumptions are as follows (1) the initial C^{14} activity in the

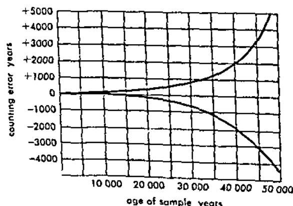


Fig 3 Measurement error for sensitive gas counting system plotted as function of sample age

Table 3 Examples of C^{14} fractionation during sample formation and the use of C^{13} measurements to correct for these deviations

Materials receiving carbon	δC^{14}*	δC^{13}†	ΔC^{14}‡
From waters of Pyramid Lake			
Tufa	−3 6	+0 63	−4 9
Algae	−8 8	−2 27	−4 3
Plants	−11 0	−2 63	−5 7
Fish	−5 9	−0 59	−4 7
Dissolved HCO_3^-	−9 3	−2 23	−4 8
From atmosphere			
Wood	+0 8	−2 50	+5 8
Atmospheric CO_2	+3 6	−0 90	+5 4

* Difference (%) between C^{14} concentration in sample and standard

† Difference (%) between C^{13} concentration in sample and standard

carbon of the unknown sample can be accurately predicted, and (2) no postdepositional alteration of the C^{14}/C^{12} ratio in the sample takes place except by radioactive decay

Deviation in radiocarbon concentration The first of these assumptions implies that the radiocarbon concentration in carbon from any material forming in the present dynamic reservoir can be precisely predicted This is not the case, however, for even though the C^{14}/C^{12} ratio in contemporary materials is reasonably uniform, significant deviations are still present These deviations are caused by three different processes (1) isotope fractionation between various carbon compounds which make up the reservoir, (2) backlogging of newly produced C^{14} in the atmosphere and surface ocean water due to the finite mixing rates, and (3) dilution of local reservoirs with "old carbon" released from the dormant carbon reservoir (mainly limestone)

Of these, isotope fractionation is the least serious since its magnitude can be accurately estimated by determining the ratio of C^{13} to C^{12} in the sample The C^{14} fractionation will be twice the C^{13} fractionation This has been verified by numerous measurements on different materials receiving their carbon from the same reservoir Examples are given in Table 3

The finite removal rate of newly formed C^{14} atoms from the atmosphere into the ocean allows the atmosphere to maintain a 10% higher C^{14}/C^{12} ratio than that in average Atlantic Ocean water The slow mixing of the ocean itself yields geographic variations in surface ocean water and maintains a higher ratio in surface than in subsurface waters The uncertainty in the magnitude of these differences introduces uncertainties in age determinations, on recent (<5000 years) oceanic materials, which are considerably larger than the experimental error

Fresh water deposits (Table 4) are greatly affected by the third process Solution of old carbonate minerals by these systems can significantly reduce their C^{14}/C^{12} ratio Although exchange with atmospheric CO_2 tends to reduce this effect, the rate is not sufficient in most cases As shown by the examples in Table 4, rivers can deviate by as much as 20% from the atmospheric C^{14}/C^{12} ratio.

The study of the variations in the contemporary *distribution of C^{14}* has been complicated by two man made processes H E. Suess has shown that the CO_2 released by the consumption of fossil fuels has measurably diluted the C^{14} in the atmosphere *reservoir* Plants growing in 1890 had an average of 2% more C^{14} than those growing in 1950. The explosion of thermonuclear devices created sufficient C^{14} to raise the atmospheric concentration in the Northern Hemisphere by 25% and in the Southern Hemisphere by 18% between 1954 and mid-1959 For radiocarbon dating much of the problem introduced by these changes is avoided by comparing unknowns with a material grown at a known time prior to 1900 Tree rings grown in 1890 are used by many laboratories *See* DENDROCHRONOLOGY.

The uncertainty in the present distribution of C^{14} introduces an error of about 100 years in age determinations performed on terrestrial plant material, of about 200 years on marine shells, and of *about 1000* years on random samples deposited from fresh water systems Further studies of the distribution of C^{14} in the various reservoirs may greatly reduce these uncertainties in many cases.

Beyond the uncertainties introduced by variations in the assay of contemporary material the possibility of time variations must be considered. *Variations* in the primary cosmic ray flux would produce corresponding changes in production rate of C^{14} and hence the total amount of C^{14} present on *the earth* Similar variations would occur if the strength of the earth's magnetic field were to *change* Since many low energy cosmic rays are deflected by the earth's magnetic field a significant *reduction* in magnetic field would result in increased C^{14} production Variations in the mixing rates of the ocean could also affect the concentration of C^{14} in the atmospheric and surface ocean reservoirs H DeVries has shown that a small

Table 4 C^{14}/C^{12} ratios for contemporary samples from fresh water systems

Laboratory	Material	Locality	ΔC^{14}*
Yale University	Several types of materials	Quecchy Lake, N Y	~ −20†
Yale University	Several types of materials	Lake Zoar Conn	~ −10†
Columbia University	Fish	Walker Lake Nev	−0 3
Columbia University	Shell	Truckee River Truckee Calif	−0 9
Columbia University	Plants	Bear River Woodruff Utah	−15 0
Columbia University	Dissolved HCO_3	Mono Lake Calif	−15 4

* Difference (%)
CO_2 over water bo[illegible]
the same C^{13} conce[illegible]
duced by fractionation

† Average of a number of results

Table 5 *Compar son* of radiocarbon ages with historic and tree r ng ages

	Un versity laboratory	H storic or *tree ring* ages	Rad o-carbon ages
Mammal an rema ns from midden at Inca *temple*	Columbia	444 ± 25	450 ± 150
Sequoia tree r ng	Columbia	880 ± 15	930 ± 100
Douglas-fir *tree r ng*	Chicago	1372 ± 50	1100 ± 150
Sequoia tree r ng	Columbia	1377 + 4	1430 ± 150
Wood from Roman sh p	*Rome*	1990 ± 3	₋030 ± ⁷00
Wood from Egyptian mummy coffin	*Ch cago*	2280	2190 ± 450
Charcoal from Etruscan tomb	Rome	*2600 ± 100*	2730 ± 240
House beam Tay nat, Syria	Ch cago	*2625 ± 50*	2531 ± 150
Redwood tree ring	Ch cago	2928 ± 52	3005 ± 165
Wood from funeral sh p n tomb of Egypt s King Sesostris	Ch cago	3750	3621 ± 180
ood *from* Egyptian tomb of Sneferu	Ch cago	4575 ± 75	4817 ± 240
ood *from* Egypt an tomb of Zoser	*Ch cago*	4650 ± 75	3979 ± 350
ood *from* Egypt an tomb of Hemaka	*Chicago*	4900 ± 200	4883 ± 200

yclic variation in the C^{14}/C^{12} concentration in the tmosphere has occurred over the past few hundred ears Since this variation correlates with tempera ure the conclusion is reached that the variation is produced by a changing rate of deep ocean water formation (more deep water being formed during cold than during warm periods)

Although the data summarized in Table 5 suggests that C^{14} ages are accurate back to the beginning of historic times no independent check exists beyond 5000 years It is generally concluded however that deviations caused by variations in production and mixing rates will be small Also since *much of the value of the method is its use in cor* relating events which occurred in different parts of the earth small changes in the absolute dates would not cause any serious changes in conclusions based *on the uncorrected data*

Postdepositional alteration There are three possibilities for postdepositional alteration of the C^{14}/C^{12} ratio in the sample (1) preferential re-

trusion of *extraneous carbon* Of these only the th rd is serious in most cases Fractionation is small and a correction can be applied by making C^{13}/C^{12} stable isotope ratio determinations Exchange is probably negligible *for organic* materials and is serious only for carbonates with very high surface areas

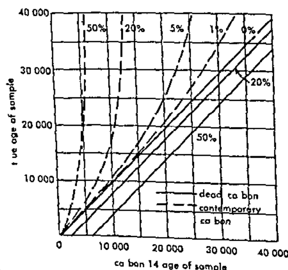
Contamination w th recent carbon presents a serious problem in dating very old samples. As can be seen in Fig 4 the addition of 1% recent carbon to a sample 40 000 years old would give it an apparent age of about 33 000 years Such contamination can occur in organic materials by *physical* in trusion (rootlets) or by precipitation of humic ma

terials leached from overlying soils *In the case of* carbonates precipitation *of* ground water carbonate *is the most likely* process

In the absence of intercalibrations with other re liable absolute methods internal crosschecks must be made Ages on various materials *from the same* horizon or on *various chemical fractions* of the *same material* provide the necessary information T types of checks have been made (

samples have been compared with those the interior of the same sample (3) the ages obtained for the base soluble fraction (which should contain any precipitated contaminants *derived from organic materials*) have been compared with those obtained on the insoluble fraction, (4) the ages for the cel lulose fraction of organic materials have been com pared with those on the lignin fraction The results are summarized *in Table 6 In general* the agreement is excellent suggesting that the contamination levels are quite low and only rarely do they lead to significant errors for samples with ages less than 40 000 years Extension of the method to sam ples *greater than* 40 000 years in age, however will require careful sample selection and extensive chemical pretreatment. Contamination will un doubtedly place an upper limit on the extension of the method *Reliable* C^{14} *ages are not expected for* samples greater than 70 000 years regardless of experimental advances

Summary The uncertainty in the assignment of the initial C^{14} concentration in the sample restricts the accuracy possible *on the absolute age of young* specimens (±100 years for terrestrial plants, ±200 years for marine shells and ±1000 years for ran dom fresh water deposits) The contamination *problem limits the extension of the method to very*



F g 4 Effect of contam nat on of samples of various ages w th varying amounts of contemporary and of anc ent ca bon.

Table 6 Internal checks of C^{14} ages

Organic-carbonate comparisons			
Laboratory	Material	Locality	Age*
U S Geological Survey	Plant remains	Utah soil	8330 ± 300
U S Geological Survey	Shell	Utah soil	7720 ± 300
Columbia University	Wood	Vancouver delta sands	11 850 ± 250
Columbia University	Shells	Vancouver delta sands	12 000 ± 250
Copenhagen	Wood	Denmark lake deposits	10 890 ± 240
Copenhagen	Marl	Denmark lake deposits	10 930 ± 300
Columbia University	Dispersed organic	Great Salt Lake sediments	26 300 ± 1100
Columbia University	Dispersed $CaCO_3$	Great Salt Lake sediments	25 300 ± 1000
U S Geological Survey	Wood	Searles Lake core	26 700 ± 2000
U S G	"	"	23 000 ± 1400
U S G	"		29 500 ± 200

Surface-interior $CaCO_3$ comparison			
Laboratory	Material	Age surface	Age interior
Columbia University	Tufa	9550 ± 250	9450 ± 250
Columbia University	Tufa	12 200 ± 300	13 000 ± 400
Columbia University	Tufa†	8800 ± 200	10 700 ± 400

Base soluble–base insoluble comparison			
Laboratory	Material	Age soluble fraction	Age insoluble fraction
Columbia University	Peat	4700 ± 150	4650 ± 150
Columbia University	Peat	8350 ± 200	7350 ± 650
Columbia University	Lignitized wood	39 000 ± 2000	39 000 ± 2600

Lignin-cellulose comparisons			
Laboratory	Material	Age lignin	Age cellulose
Columbia University	Wood	25 850 ± 500	25 900 ± 300
Columbia University	Peat	25 050 ± 300	23 450 ± 300

* Years before present
† Very high surface area

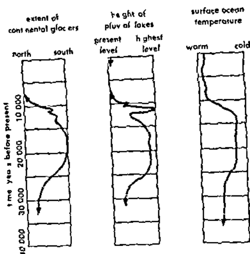
old samples (organic materials to <60 000 years and inorganic <40 000 years) In the intermediate range the experimental error dominates in most cases

APPLICATIONS OF RADIOCARBON DATING

The radiocarbon method has been applied to numerous problems Perhaps the most important is that of establishing the chronology of the climatic changes which characterize the Pleistocene period Carbon 14 ages of samples from trees knocked over by advancing ice masses record the chronology of advancing continental glaciers Peat samples from bogs and driftwood from the shorelines of pro glacial lakes make it possible to establish a time table for the retreat of glaciers Carbonate samples from the shorelines of the great pluvial lakes (which once covered many of the desert areas in the Great Basin of western United States) allow the absolute history of the climate fluctuations which produced these changes to be established Radiocarbon dates on the shells of planktonic ani mals found in deep sea sediments define the chro nology of fluctuations in oceanic conditions and re- lated climatic conditions (*see* MARINE SEDIMENTS) Oxygen isotope measurements on the same shells permit a quantitative estimate to be made of sur face ocean water temperature at a given time (see

past 40 000 years As shown in Fig 5 in each case the data suggests a warm climate during the past 10 000 years preceded by a cold period extending back to about 27 000 years Prior to 27 000 years conditions were intermediate between those char acterizing the interglacial climate of the present and the glacial climate for the period preceding 10 000 years ago

Radiocarbon dates on charcoal from the hearths of ancient man have been a great aid in working out man s history and relating it to the climatic fluctuations mentioned above Volcanic eruptions have been dated by radiocarbon measurements on materials covered by the lava or ash, age determi ations on hydrocarbons extracted from soils yield valuable information concerning the origin of pe troleum dates on the remains of extinct animals such as the mastodon and giant sloth allow esti mates to be made of the time and cause of extinc tion and dates on charcoal and artifacts left by modern man in remote areas allow much



5 Climate chronologies obtained from carbon measurements on samples from continental glacier deposits pluvial lake deposits, and ocean sediments right-hand side of each graph represents cool moist conditions of periods of glaciation

history to be related to that of cultures which have provided historic records

Carbon 14 dates from all laboratories are now being continuously compiled on a punch card system by Radiocarbon Dates Association, Inc Andover Mass [W S B]

Bibliography W F Libby *Radiocarbon Dating*, 2d ed 1955.

Radiochemical laboratory

An installation used to provide a safe environment for handling and investigating radioactive materials Common features of radiochemical laboratories include the use of readily decontaminated surfaces good ventilation cleanliness good lighting special arrangements for waste disposal, and filtration of both supply and exhaust air Specific features of these laboratories depend primarily on the type and amount of radioactivity handled

If only α-emitting isotopes are involved protection of personnel and prevention of contamination are provided by use of enclosed hoods or boxes equipped with rubber gloved openings (Fig 1) For work with very small quantities of α emitting isotopes the use of gloves on the openings is sometimes dispensed with In this case light rubber hand gloves are worn by personnel and special provisions are made to control the size of hood openings and to maintain a substantial negative

ferential serves to sweep air away from the operator into the hood so that dangerous vapors in the hood will not be inhaled An air velocity of about 125 ft/min through hood openings is regarded necessary to achieve this safety

For work with β and γ emitting isotopes, more shielding is required than is obtainable with rubber gloves and structural material of normal thickness in laboratory hoods When the radioactivity is of moderate amount and conveniently confined, adequate shielding may be obtained by lead or other metal blocks placed around the radioactive source To avoid overexposure to hands, long handled tongs with suitable head design are used to perform various manipulations When the radiation is more intense and less confined it is necessary to depart from essentially direct contact operation to remote operation In addition, the working area must be enclosed with thick walls of shielding material

Such facilities, called hot laboratories, are a special type of radiochemical laboratory and are usually referred to as caves or hot cells They are usually provided with remote manipulators and thick viewing windows Caves with shielding limited to handle 1–10 curies of γ radioactivity are generally called junior caves (Fig 2)

Ventilation. One of the most important items required for radiochemical laboratories is filters for inlet and exhaust air Because particulate matter such as dust is a good vehicle for spread of radioactivity, it is desirable to provide clean air supply by inlet filters Clean air thus supplied helps to reduce the dust loading on exhaust filters It is also extremely important to filter exhaust air from radiochemical laboratories This is especially so when large amounts of plutonium and β γ activities (as in caves) are handled Very effective filters are employed for this purpose in order to avoid contamination of the environment

Because work with radioactivity always includes risk of contamination, radiochemical laboratories include substantial areas for clothing change, showers, clothing storage, and lockers Working area surfaces including floors hoods, and caves should be smooth and as continuous as possible to permit relatively easy decontamination Structural materials and paints should be selected to permit effective decontamination procedures

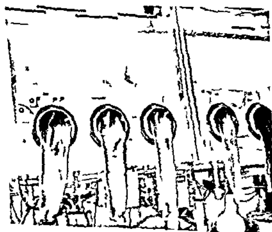


Fig 1 Radiochemical laboratory, hood equipped, with rubber-gloved openings (Argonne National Laboratory)

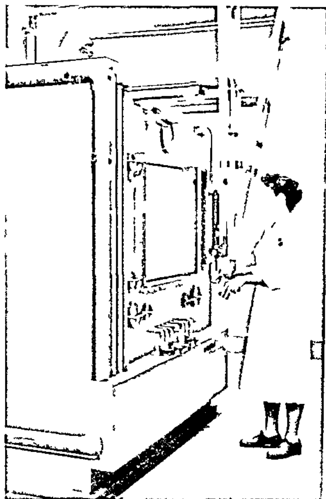


Fig 2 Junior cave operating face and end wall (shielding including window is equivalent to about 1 ft of dense concrete) (*Argonne National Laboratory*)

Operations In general the variety of work carried out in radiochemical facilities corresponds to that conducted in conventional laboratories There fore there is need for similar kinds of measurements and examinations In the case of radiochemical facilities it is very often necessary to conduct these activities in hoods and caves and therefore instruments and other laboratory equipment must be located within these enclosures For these reasons it is important to select equipment especially instruments such as balances which is most suitable with the restricted manipulation and viewing available

Work conducted in caves places a premium on good visibility adequate shielding and effective remote manipulations Viewing windows are the means most frequently used for observations These are available in various thicknesses to permit as much shielding effectiveness as is incorporated in other parts of the enclosure Viewing windows must have good light transmission and minimal distortion Resistance to discoloration caused by high levels of radiation is desirable for panes located on the inner side of caves Bulk shielding may be obtained by multiple layers of glass plate or by a combination of glass and dense zinc bromide solutions (about equivalent to ordinary glass and concrete in shielding effectiveness) When very dense shielding (for example steel) is used elsewhere in the enclosure, the windows are thicker than the rest of the shielding wall unless special high-density glass is used Supplementary means for observation sometimes include periscopes and closed-circuit television The best possible lighting should be installed in caves to assure viewing without eye strain

Shielding A wide variety of shielding materials is used for caves Ordinary concrete is frequently selected because of its relatively low cost When limited floor area is available denser concrete is employed Increased density is obtained by use of barytes in the concrete formulation or by addition of iron shot For still more effective shielding, magnetite steel and lead are used Steel and lead are often preferred for special purposes such as movable partitions and closures for material transfer ports Whatever kinds of shielding materials are employed in cave construction the final structure should be checked thoroughly with the aid of appropriate radiation sources and detection instruments This will reveal whether or not there are cracks or voids which might represent regions of deficient shielding

Remote manipulation The design construction, and installation of remote manipulators have received considerable attention A substantial variety of these devices is available commercially The most frequently used type is the mechanical master slave manipulator originally developed by the Argonne National Laboratory It is a general purpose manipulator most suitable for junior-cave (Fig 2) and intermediate-level shielded enclosures It penetrates the shielded enclosure through a hole near or through the roof This penetration is an undesirable feature of this type of manipulator because it cannot be completely sealed and therefore introduces a risk of scatter radiation The manipulator can be booted for contamination control (Fig 3) and has remotely removable tongs



Fig 3 Interior view of a high level radiochemistry cave (shielded for 10 000 curies of γ-activity) (*Argonne National Laboratory*)

Its load capacity is about 10 lb in any direction An important feature of these manipulators is that there is about a 1 1 correspondence between the force applied by the operator and that applied by the slave end in the cave. Manipulators with many independent and precisely controlled motions and ... have been designed and built

...ions available with mechanical mas... ...nipulators, but also the electrical linkage in ... of mechanical linkage makes possible tighter ...ls at penetrations through the shielding walls ... more flexible locations of the master slave end ...ative to the operator s location

...afety The operation of radiochemical labora...ies requires persistent monitoring of radioac...ity to assure safety to personnel and control of ...ntamination to the environment In addition ...ergency power should be installed to guarantee ...ntilation of critical areas if disruption of regu... power should occur The ventilation should be ...pable of maintaining negative pressures relative ... the outside face for any enclosure (hoods or ...ves) containing substantial levels of radioactiv... Failure to do this may result in serious conse...uences because the radioactivity may diffuse out...de the enclosure into areas normally occupied by ...ersonnel Another important consideration for ra...iochemical laboratories is adequate provision for ...he handling and disposal of various kinds of radio...ctive wastes

for work areas such as high level caves See NUCLEAR FUELS REPROCESSING RADIATION SHIELDING RADIOCHEMISTRY [S L]

Bibliography *Proceedings of the Seventh Hot Laboratories and Equipment Conference 1959 Nuclear Congress Cleveland Ohio* 1959 U S Atomic Energy Commission *Chemical Processing and Equipment* 1955

Radiochemistry

A subject which embraces all applications of radioactive isotopes to chemistry The subject is not precisely defined and is closely linked to nuclear chemistry The widespread use of isotopes in chemistry is based on two fundamental properties exhibited by all radioactive substances The first property is that the disintegration rate of an isotopic sample is directly proportional to the number of radioactive atoms in the sample Thus measurement of its disintegration rate (with a Geiger counter for example) serves to analyze a radioactive compound With nearly all chemical elements (the most notable exceptions being nitrogen and oxygen which have no suitable radioactive isotopes) an isotope may be incorporated in a chemical compound and thereafter masses of the compound as small as 10^{-6} 10^{-10} g may be measured with a high precision Because experimental chemistry depends largely upon analysis isotopes may be employed in most chemical problems especially those requiring high analytical sensitivity The second fundamental property is that the disintegration rate is completely unaffected by the chemical form of the isotope and conversely the property of radioactivity does not affect the chemical properties of the isotope By substituting or labeling a particular atom within a molecule isotopes can be used to trace the fate of that atom during a chemical reaction However in contrast to physical migration tracer studies the various compounds arising in a reaction must first be isolated in separated pure forms before radioactive assays can be performed See TRACER RADIOACTIVE

In general radiochemical studies can be classified according to whether the use of isotopes represents a convenient or a unique solution to a problem

Convenient applications These applications usually exploit the high sensitivity of tracer techniques because alternative analytical procedures are slower and often less accurate The efficiencies of chemical separations such as those based on selective precipitation solvent extraction ion exchange and electrodeposition reactions, are studied by labeling the desired compound and following the radioactivity during the separations The rate and extent of adsorption on solid surfaces of either labeled solutions or labeled gases are rapidly determined by assaying periodically the mobile phase or better the solid phase New chemical phenomena such as the coprecipitation of trace elements and radiocolloid formation occur at submicro concentrations (10^{-10} g/liter of solution) and may be studied most conveniently with isotopes The solubility of an insoluble precipitate is measured by saturating a solvent medium with a radioactive solid Similarly the vapor pressure of a solid is measured by saturating an evacuated volume with vapor or for pressures below 10^{-4} mm by effusing vapor onto a cooled target which is later assayed Qualitative and quantitative analysis for most trace elements present in parts per million or less in a sample is possible by radioactivation analysis The sample is irradiated in a flux of neutrons or other suitable particles and the trace element is identified and determined by its induced activity Depending upon the element quantitative determination of masses of 10^{-6} 10^{-12} g is usually possible See NUCLEAR REACTION

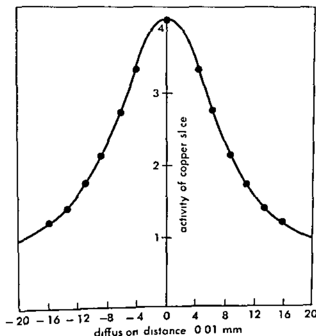
Unique applications An understanding of diffusion processes is of considerable importance because the rates of many chemical reactions are governed by the rate at which chemical species can diffuse through a medium to the point of reaction For example the rate of many electrode processes depends upon the rate of diffusion of electrolyte to the electrode and the rate of oxidation of copper is determined by the rate of diffusion of

copper ions up to the metal surface If a layer of radioactive copper is sandwiched between two ordinary copper samples, it is found that at elevated temperatures copper ions will diffuse considerable distances within the metal The rate of diffusion of copper in copper (that is, the self diffusion rate of copper) can be observed only by the transfer of radioactivity from the labeled region into the unlabeled regions, thin slices being removed from the solid at known distances and then being assayed The illustration shows some experimental points obtained in such an experiment The distribution curve is that expected from the integrated form of Fick's law of diffusion,

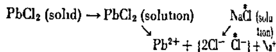
$$\partial c/\partial t = D(\partial^2 c/\partial x^2)$$

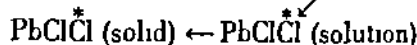
where c is the concentration (activity) of the diffusing tracer, t is the diffusion time, x the diffusion distance, and D the self diffusion coefficient In addition to solid state studies with elements, alloys metallic oxides, and inorganic salts (all of which have important metallurgical implications), self diffusion experiments are performed with liquids and gases In all cases, they provide valuable information on the nature of the intermolecular forces which determine the magnitude of D

Isotopic exchange reactions When slightly soluble lead chloride crystals are mixed with an aqueous solution of sodium chloride, labeled with chlorine 36, radioactivity rapidly appears in the lead chloride as a result of exchange of chloride ions between the two compounds Both compounds produce chloride ions on dissociation, and some chloride ions, originating from the sodium chloride, become associated with the lead ions, thereby leading to radioactivity in the lead chloride

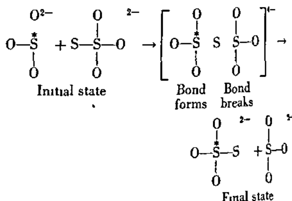


Distribution of radioactive copper after 605 min diffusion at 950°C The full curve corresponds to $D = 8.92 \times 10^{-10}$ cm²/sec (*Calculated from the results of M S Maier and H R Nelson, Self diffusion of copper, AIME Trans, 147 39–47, 1942*)

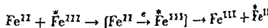




Exchange processes are proceeding continuously but they can be detected only with isotopes, hence the term isotopic exchange reactions Exchange reactions may occur between any two species of molecules having a common atom or group They may be due to a dissociation process (as above) or to a collision between the two species in which chemical bonds are formed and broken (a bimolecular process), as in the exchange of radioactive sulfur (denoted by an asterisk) between sulfite and thiosulfate ions



With two oxidation states of an element exchange occurs by the transfer of an electron (an electron exchange reaction) for example



although in many cases, the electron may actually be transferred by an atom or a group bridging between the reactants The rates of exchange reactions are measured by separating the two reactants at various times and determining the fraction F of the total radioactivity in each species For an ini

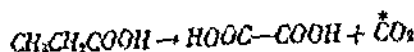
in each species for an initially labeled species

upon reactant concentrations Rate studies of ex

with molecules (solvolysis) the thermal dissociation of gases and the dissociation of gases on catalyst surfaces Bimolecular exchanges are exceptionally important in studies of oxidation reduction reactions and of substitution reactions of coordination complexes of the transitional elements The rates of exchange by substitution provide

a rect measure of the lability of these coordination complexes

Isotopic tracer studies Details of reaction mechanisms are provided by labeling a specific atom within a molecule Thus when carboxyl labeled propionic acid is oxidized in acid dichromate the product as anticipated is only radioactive carbon dioxide



$$CH_3CH_2COOH \rightarrow HOOC{-}COOH + \overset{*}{C}O_2$$

However the oxidation mechanism is more complex in the case of alkaline permanganate because the isotopic distribution is 75% labeled oxalate and only 25% carbonate When the masses of the labeling and the normal isotopes differ markedly (notably with hydrogen carbon and oxygen) the isotopic molecule will react more slowly than the normal molecule if the reaction mechanism involves significant stretching of the chemical bond to the isotopically substituted atom Such isotope effect studies thereby provide further mechanistic information

New elements The artificially produced transuranium elements and technetium astatine and francium are all radioactive and their chemistry is being elucidated with radiochemical techniques However this topic is normally accepted as nuclear chemistry

Recoil studies For a discussion of recoil studies see NUCLEAR CHEMISTRY See also RADIATION CHEMISTRY RADIOACTIVITY RADIOCHEMICAL LABORATORY [D R S]

Bibliography G Friedlander and J W Kennedy *Nuclear and Radiochemistry* 1955 A C Wahl and N Bonner (eds) *Radioactivity Applied to Chemistry* 1951

Radio frequency amplifier

A tuned amplifier amplifying signals in the radio-frequency (rf) region The frequency of the maxi

shunt capacitance which adversely affects the gain of an RC coupled amplifier forms a part of the tuning capacitor in the rf amplifier allowing a high gain per stage See AMPLIFIER RADIO RECEIVER [H F K]

Radio frequency spectroscopy

The branch of spectroscopy concerned with the measurement of the intervals between atomic or molecular energy levels that are separated by frequencies from about 100 kc to 1000 Mc (10^5 10^9 sec^{-1}) as compared to the frequencies that separate optical energy levels of about 6×10^{14} sec^{-1} The importance of radio-frequency spectroscopy lies in the fact that certain specific properties of the nucleus such as spin magnetic dipole moment and electric quadrupole moment play a relatively major role in determining the intervals between closely lying energy levels the results of this branch of spectroscopy have been of great importance in determining nuclear properties For a discussion of radio frequency spectroscopy by what is called the molecular beam magnetic resonance method see MOLECULAR BEAMS, see also MAGNETIC RESONANCE MICROWAVE SPECTROSCOPY NUCLEAR MOMENTS SPECTROSCOPY [P KU]

Radiography

T... of making photoshadowgraphs of ... object it is differentially absorbed depending upon the va... composition ... the emerge... to the specimen under examination constitutes a shadowgraph or radiograph of the interior of the specimen Radiography is the general term applied to this nondestructive technique of testing the gross internal structure of any object, whether it be of the chest of a human patient for evidences of tuberculosis silicosis heart pathology or imbedded foreign objects of a metal casting suspected of possessing internal cracks or other defects or of a multitude of other objects whose inner structures are unrevealed by visual inspection

The term roentgenography specifically applies to the use of x radiation especially in medical diagnosis Microradiography is simply the extension of radiography to small specimens and the enlargement of images When the image is registered on a fluorescent screen and visually observed the technique is called fluoroscopy The photography of the fluorescent image, as in mass chest examinations is called photofluoroscopy The radiation sensitive silver halide emulsion of photographic film may be replaced by the dry plate electrostatically formed image of xerography and the technique becomes xeroradiography See MICRORADIOGRAPHY RADIOLOGY

The wide variation in absorbing power of various substances for x rays was recognized by W C Rontgen in his first experiments following his discovery of x rays in 1895 The radiographs of his wife s hand showing bones and wedding ring, of the weights in a closed box and the interiors of other objects are among the classics of science See X RAY(S) PHYSICAL NATURE OF

Among the many objects now examined by radiography are the human chest human teeth metal castings coal minerals rubber tires, golf balls fabricated objects with internal seals electrical equipment bottles and containers of all kinds grain fruit meats battery plates and paintings The use of radiography in the field of metallurgy

is covered in a separate article (*see* RADIOGRAPHY OF METALS)

Absorption laws Radiography is a straightforward application of the well known exponential absorption law

$$I_x = I_0 e^{-\mu x} = I_0 e^{-(\mu/\rho)\rho x}$$

where I_0 is the initial intensity of the x ray beam, I_x the intensity after passage through the object of thickness x ρ the density μ the linear absorption coefficient and μ/ρ the mass absorption coefficient The ratio μ/ρ is a function of the atomic number of the absorbing chemical element and of the wavelength of the x ray beam The value of μ/ρ for a compound is equal to the sum of the values for each of the elements present The penetrating ability of x rays increases as the wavelength of the rays decreases (hard x rays) The x rays of larger wavelength (soft x rays) contain relatively less energy The coefficient μ/ρ is proportional to the cube of the wavelength The product ρx is mass per unit area

This exponential law and related expressions quantitatively predict why bones absorb more radiation in passage than soft tissues and thus are clearly delineated or why cracks or blowholes in castings absorb to a lesser degree than the solid metal and thus are disclosed on the radiograph

Equipment Because of the wide range of materials subjected to radiographic test a wide range of operating voltages for x ray tubes is required The higher the tube voltage the greater the energy and penetrating power (or the lower the wavelength) of the x ray beam (*see* X RAY TUBE) Somewhat arbitrarily the equipment and operating conditions may be summarized in the manner shown in the preceding table

Types of x ray tubes used in radiography

Energy of radiation	X ray tube	Power source	Application
1 50 kev	Conventional	High tension transformer	Small parts, thin foils, dental radiography
50 250 kev	Conventional	Regular or resonance transformer gas filled portable with tube integral	Usual range for medical radiography or industrial objects up to 4 in of metal
250–500 kev	Conventional	Cascade or resonance transformer a few γ emitting isotopes	Heavier metal sections
1–2 Mev	Sealed tube with high potential applied over gradient in center of high potential generator	Resonance transformer or Van de Graaff Generator with tube integral γ-emitting isotopes such as cobalt-60	Metal sections up to 12 in
10–24 Mev	Donut	Betatron	Sections up to 20 in or objects of greatly different cross section

General principles and techniques Some of the more important principles and techniques common to radiography in general are given in the following paragraphs

Photographic film Since radiographs are usually interpreted from photographic negatives, each object under examination must be subjected to a controlled technique and properly selected film in terms of contrast, latitude and sensitivity Optimum blackening is $S = 0\,7–0\,9$ where $S = \log L_0/L$ The photometrically measured light intensities are L_0 before and L after the passage through the photographic layer The normal eye can detect with certainty a minimum blackening difference between adjacent areas of 0 02

Contrast is the difference in density produced by a change in object thickness Latitude is the extent of object thickness that can be reproduced in the working range of blackening S Sensitivity is the smallest fractional increase in thickness detectable, the practical limit is 0 02 Definition is the fidelity (or sharpness) with which a radiograph delineates a discontinuity Speed is the characteristic relative to initial period of inertia in exposure, and is controlled by intensifying screens *See* PHOTOGRAPHY

Target to specimen distance This should be as great as possible (usually 20–30 in) consistent with the fact that radiation intensity decreases inversely as the square of the distance in order to decrease the unsharpness of images caused by penumbra (Fig 1)

Focal spot size This should be as small as possible to gain emission from (as nearly as possible) a point source since otherwise there is unsharpness

cause fogging and poor definition they are controlled by lead screens or diaphragms consisting of a [illegible]

which

sition

sure

Voltages and exposure charts For each type of object to be radiographically examined there is an optimum range of voltages controlling penetration and times of exposure The exposure values (log of milliampere seconds) for each voltage are plotted against the thickness of specimens of particular composition

Avoiding over and underexposure Equalizing of

(iodine barium and thorium compounds) is a common procedure for soft tissues in medical di-

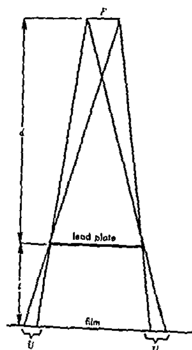


Fig 1 Geometry of radiographic image formation F size of focal spot of x-ray tube d, distance from focal spot to specimen (plate representing an internal defect), t distance from defect to film $U = Ft/d$, additional width of penumbral shadow (or unsharpness)

of a given material as a function of radiation energy flattens out over a considerable range of energies and wavelengths with the result that thin and thick sections are correctly exposed The automatic pistol Fig 2 is an example

Special techniques A number of special radiographic techniques have been developed primarily for medical diagnosis but are also applicable to industrial testing Some of these are now listed

Stereoscopy Two exposures are made at tube positions roughly corresponding to interpupillary distances (2⅝ in) The two negatives are observed in a stereoscope which fuses the two images into one with the illusion of a third dimension This permits location of any detail of gross structure (defect in a casting bullet in a chest, location of cancer) below the surface

Tomography Of the three components—tube, subject and film—two are moved during exposure synchronously so that it is possible to register the radiographic image of one plane in the object while images outside this slice have a continuous relative displacement and are blurred A series in various planes enables three-dimensional exploration called serioscopy

Kymography Internal motion in an object (heart beat breathing of lungs) can be recorded by interposing between ... ect and film a lead diaphragm with a 1 mm slit The film is moved at a speed compatible with the motion and records a saw tooth image as a time record of such motion

Instantaneous radiography With an intense x ray beam from field emission pulse discharge, it is possible to make exposures in 10^{-6} sec, so that a rapidly moving object such as a bullet or shell appears to stand still This has been of great significance in studies of ballistics moving parts in engines, and other objects A series of high intensity exposures leads to cineradiography—moving pictures Exposures of 01 001 sec are conveniently made with rotating target x ray tubes, which permit passage of electron currents of 500–1000 milliamperes without damage to the target. Thus, images of heart and lungs or moving machinery may be registered between successive motions (Fig 3)

Monochromatic radiography By proper techniques of filtration, or reflection from crystal faces or mirrors, it is possible to produce monoenergetic beams, especially below 50 kev The radiographs of thin or small objects (especially when magnified) are greatly superior to those from beams with a whole spectrum of rays each with a different absorption in the specimen

Electronic radiography This depends upon the emission from irradiated specimens of electrons whose energies and photographic effects are highly characteristic of the specimen A beam of hard x rays passes through a film without affecting it onto a specimen from which electrons are liberated backward to form the latent image, or the beam ejects from a metal foil electrons which pass

graphs by radio or telephone is now an accomplished fact Elements of the original negative are transmitted exactly as are news photos and the receiver re-records the whole image This is of special value in medical or industrial diagnosis at a dis-



Fig 2 Radiograph of automatic pistol made with 10-Mev x rays generated in a betatron, showing thin and thick sections properly exposed because of absolute sensitivity of radiation

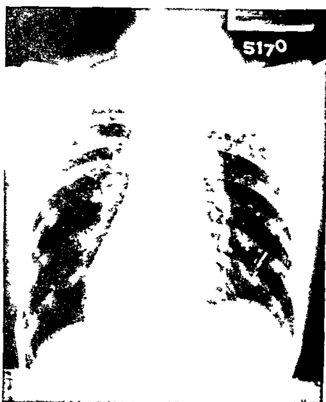


Fig 3 Chest radiograph of foundry worker made with intense beam from rotating target x ray tube, showing nodules in lungs due to silicosis and shadows of skeleton, heart, and stomach

tance by a central agency or by an expert whose immediate interpretation is required

New developments Among the new techniques being developed are (1) the use of color film where far more sensitive differences are delineated (2) completely automatic procedures from start to finish, including evaluations, with technicians entirely remote and of course subjected to no radiation exposure [G L CL]

Bibliography G L Clark, *Applied X rays*, 4th ed , 1955

Radiography of metals

A method of nondestructive inspection of metals which detects the internal physical imperfections such as voids, cracks, flaws, segregations, porosities and inclusions It guarantees the quality of the material to the customer and contributes to improved production processes when expertly coordinated It visualizes the actual location and physical condition of otherwise unaccessible parts within an object

The penetrating characteristics of x rays and γ rays are used The sources of such radiation are x ray generators, linear accelerators, betatrons, and radioisotopes that have energies ranging from about 10 kv to 24 Mev

The thickness, physical density, and absorption coefficient of the metal under investigation determine the most suitable energy of radiation (Fig 1)

The recording media most commonly used are double-coated films sensitive to electromagnetic radiation, fluoroscopic screens, or electronic devices.

Radiography is extensively used wherever internally sound metallic components are required such as (1) in the foundry industry to guarantee the soundness of castings, (2) in the welding of pressure vessels, pipe lines ships and reactor components, to guarantee the soundness of welds (3) in the manufacture of fuel elements for reactors, to guarantee their size and soundness (4) in the solid propellant and high explosives industry to guarantee the soundness and physical purity of the material, and (5) in the automotive, airplane, atomic, and guided missile industries, whenever internal soundness is required

Film radiography. For better interpretation of the size, shape, and location of defects, two or more radiographic views of the object are obtained under predetermined angles to each other Stereoradiography is also used for this purpose

If the thickness of the specimen varies, complete coverage can be obtained by one of three methods (1) various radiographs may be taken under exposure conditions as determined by the average thickness of the respective section of the object (2) the multiple film method may be applied by placing a combination of fast and slow films in the same film holder, and (3) compensators may be used by building up the thinnest parts of the object to an average uniform thickness with similar radiographically opaque material

Radiographic inspection. The adequacy of the radiographic technique is determined by the use of penetrameters They usually are plates of the same metal as the object under investigation and their

can be seen in the radiograph the penetrameter sensitivity is 2% In certain phases of industrial radiography the term radiographic resolution is used when expressing the quality of the radiograph In these cases a series of penetrameters is

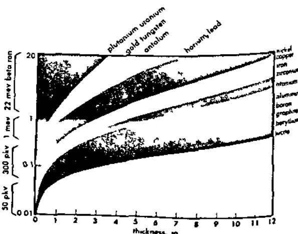


Fig 1 Relation of the penetration characteristics of x rays to the type and thickness of various materials

sed of various thicknesses, having holes with diameters equivalent to their thickness The diameter of the smallest radiographically detectable hole expressed in millimeters or inches determines the respective radiographic resolution Radiographic inspection procedures as accepted by private and governmental industry or technical organizations, require the use of such penetrameters (Fig. 2)

Fluoroscopy. This technique permits the examination of the object by direct observation of certain fluorescent screens when exposed to the electromagnetic radiation transmitted through the object

The primary advantage over the film method is its economy, the primary disadvantage its inferior sensitivity

An optically transparent but radiographically opaque barrier such as lead oxide glass is placed in front of the screen to protect observers from radiation Because of this hazard direct fluoroscopy is used up to an energy of about 150 kv only and it is mainly applied to the inspection of light metals or light alloys such as aluminum and magnesium castings of thicknesses up to almost 3 in Gross defects such as gas holes shrink cavities and inclusions can be detected if their size is at least 6-8% of the total thickness of the object

Polaroid radiography. This technique includes a one step developing process through which a nearly dry paper supported radiographic image is obtained within about 2 min after exposure without the use of a conventional darkroom or wet processing This process is best suited to low energy radiography The short density range of the Polaroid paper is most applicable to radiography of high contrast specimens such as metal plastic assemblies

Electronic radiography The basis for this technique is the application of direct image converter

ng using x ray energies up to 2000 kv to inspect teel up to 3 in thick or equivalent thicknesses of other materials

It is superior to direct fluoroscopy because of increased brightness of the radiographic image

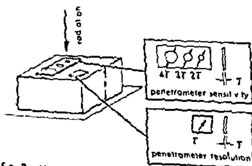


Fig 2 Use of penetrometers. *T* is the penetrometer plate thickness.

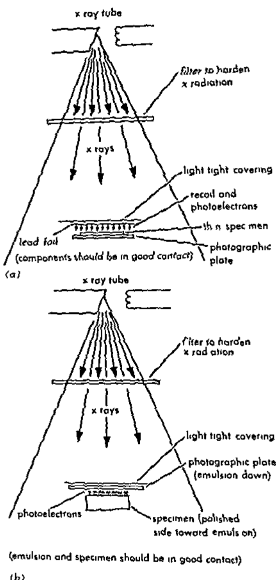


Fig 3 (a) Transmission electron radiography (b) Back emission electron radiography

It is safer because the viewing screens do not have to be in the beam of radiation

Microradiography The use of photographic enlargement of fine grain radiographs gives better evaluation of the findings in microradiography

Advantage is taken of the intense electron emission of dense material when exposed to x rays. Transmission electron radiography (Fig 3a) makes it possible to obtain radiographic images of very thin specimens, and back emission radiography (Fig 3b) visualizes among other things, the presence of material of different atomic numbers in the surface of the specimen

Limitations of radiographic inspection Limitations are determined by the radiation energy, the thickness of the material, the sensitivity of the various recording devices and the shape size, and location of the defects In many cases it is a satisfactory method but there is often a need for com

plementary, nondestructive inspection methods, such as ultrasonics, magnetic particle inspection or electromagnetic test methods *See* METAL INSPECTION, MAGNETIC, METAL INSPECTION, ULTRASONIC, MICRORADIOGRAPHY, RADIOGRAPHY

[G H T]

Radioisotope

A radioactive isotope (as distinguished from a stable isotope) of an element Atomic nuclei are of two types, unstable and stable Those in the former category are said to be radioactive, and eventually are transformed, by radioactive decay, into the latter One of the three types of radioactive ray (α, β, and γ rays) is emitted during each stage of the decay The term radioisotope is also loosely used to refer to any radioactive atomic species Whereas approximately a dozen radioisotopes are found in nature in appreciable amounts, hundreds of different radioisotopes have been artifically produced by bombarding stable nuclei with various atomic projectiles (*see* RADIOISOTOPE PRODUCTION) For a discussion of radioactive radiations and their duration, *see* RADIOACTIVITY *See also* ISOTOPE

[H E D]

Radioisotope (assay)

An analytical technique including procedures for separating and reproducibly measuring a radioactive tracer Separations for the assay may be made carrier free (without stable atoms added) or with a few milligrams of added carrier by techniques such as ion exchange, solvent extraction, distillation, precipitation, electrodeposition, and isotopic exchange Once separated, the radioisotopes may be measured as gases, liquids, or solids in Geiger Muller, proportional, or scintillation counters, or in ionization chambers Each detector has characteristic advantages and disadvantages and is used only after consideration of the amounts and types of radiations emitted by the radioisotope Weak beta emitters such as carbon 14, sulfur 35 and particularly hydrogen 3 (tritium) require special techniques to eliminate errors caused by absorption of their rays in the sample

Geiger and proportional counters are used primarily for measuring beta rays, whereas solid scintillation detectors are used for gamma rays Liquid scintillation counters eliminate many self absorption problems with weak beta emitters

Measurements are usually referred to a standard of known absolute strength When absolute assay is required, problems of geometry, back scattering self absorption, counter efficiency, and the decay scheme of the radioisotope must be considered *See* ACTIVATION ANALYSIS, PARTICLE DETECTOR, RADIOISOTOPE PRODUCTION, RADIOMETRIC ASSAY

[W ME]

Radioisotope (biology)

Radioisotopes are radioactive isotopes This article discusses their usefulness in biology as tracers in the investigation of metabolic processes

The usefulness of radioisotopes as tracers arises chiefly from three properties (1) at the molecular level the physical and chemical behavior of a radioisotope is practically identical with that of the stable isotopes of the same element, (2) radioisotopes are detectable in extremely minute concentrations; and (3) analysis for radioisotopic content often can be achieved without alteration of the sample or system In some applications radioisotopes are not essential but are used because of the advantages of sensitivity or convenience. In other applications, however, principally those in which reaction rates and transfer rates are studied, isotopes, particularly radioisotopes have unique advantages as tracers and their use in studies of the kinetics of steady state systems, self-diffusion, and metabolic pathways has opened numerous fields previously thought to be inaccessible *See* ISOTOPE RADIOISOTOPE; RADIOISOTOPE (ASSAY)

Methods. Radioisotope methods include preparing the labeled material, introducing it into the system to be studied, detecting its presence and interpreting the results

Preparation of labeled compounds is one of the most important and most difficult steps in tracer methodology In synthesis with radioactive reagents, the chemical procedures used must often be modified for remote handling, and other special precautions must be taken to prevent contamination of laboratories and personnel with radioactive material and to minimize irradiation of the personnel Containment, shielding, and distance are the main items in protecting personnel from radiation. Some compounds labeled by biosynthesis, derived from bacteria or plants grown on media containing radioactive substrates, are commercially available. Before a labeled product is used it *should be tested*

states has caused confusion The chemical stability of the label (that is the radioactive element) should also be established It is unwise to use material labeled in easily exchangeable positions if the fate of the material itself (rather than the exchange process) is the point of interest *See* RADIO-

by many factors Sufficient tracer to be detectable must be used but the amounts of material which are introduced must be small enough not to disturb the system either by their mass by pharmacological effects or by the effects of radiation The mass of 1 curie the unit of disintegration rate depends inversely upon the half life and directly upon the atomic weight of the particular radioisotope, it is 1 g for radium 226 (half life 1620 years), but only 8 micrograms for iodine 131 (half life 8 0 days) In tracer experiments with small animals, microcurie quantities are usually adequate It is sometimes necessary to ch

... isotope in a single injection or using a con... ... example in order to main...

...esulting data are easier to analyze ...
...ere are many methods for detecting the pres... of radioactive material The Geiger counter ...largely been displaced by thallium activated ...im iodide scintillation crystals for counting ...s but Geiger counters and proportional count...re still useful for counting α and β particles ... PARTICLE DETECTOR) A well type scintillation ...ter may have more than 25% efficiency for ...

In a scintillation crystal the passage of a ...y through the crystal produces a flash of ...t which is detected by a photomultiplier tube ...nected to the crystal (see PHOTOTUBE MULTI...ER) The signal from the photomultiplier tube ...y be fed into a scaler or rate meter as with the ...al from Geiger tubes or after amplification it ...y be used as the input signal for a pulse-height ...lyzer in a method called γ ray spectroscopy ...e SPECTROSCOPY) The pulse height analyzer ...kes it possible to record pulses of a single ...ight corresponding to single γ ray energies so ...t one isotope may be counted selectively in the ...esence of others With multichannel pulse-height ...alyzers simultaneous determinations may be ...ade of a number of γ emitting isotopes in a mix...re Beta particle spectroscopy is also possible but ...uch more difficult both to achieve and to in...rpret in part because β particles are emitted ...ith a spectrum of energies whereas γ rays are ...pically monoenergetic For very low energy ... particles (negative electrons) particularly those ...mitted by C^{14} and tritium (H^3) liquid scintilla...ion counting is useful (see SCINTILLATION DETEC...OR LIQUID) In this method a solution of the sam...le and a phosphor takes the place of the scintilla...ion crystal Low energy β emitters can also be ...ounted by introduction in the gas phase directly into proportional counters In histological and cy...ological studies the method of autoradiography in which photographic film is exposed through contact with the specimen is very useful The autoradio-graphic method is also used extensively in con-junction with paper or column chromatography particularly in studies of metabolic pathways See AUTORADIOGRAPHY CHROMATOGRAPHY

In many applications the interpretation of the data obtained with radioisotopes is quite simple The volume of dilution and total exchangeable mass for example are inversely related to the final concentration Following a metabolic pathway or determining the uptake of radioisotopes by vari-ous organs or tissues is analogous to following a flock of sheep by belling the leader In studies of rate processes however particularly when material is moving in opposing directions simultaneously extensive use of mathematics is needed for correct interpretation of the data The mathematical basis for such interpretation is outlined in the following paragraphs

... times of a system as ...

and out of the compartment are equal ... ary but not always realistic to assume that within a given compartment mixing of the labeled and un labeled material is instantaneous and complete, the specific activity (the ratio of the amount of the labeled form of a substance to the total substance present) within a compartment is assumed to be uniform in space, although it may be a function of time.

Given a compartmental model regarded as repre-senting the system of interest (often greatly simpli fied) the expected behavior of the tracer is described with a set of differential equations in terms of the invariants of the system. The solutions of these equations contain constants which would be determined experimentally from observations on an ideal system Although prediction of tracer behavior from information about the system has many applications, it is the reverse problem that of using the experimentally determinable constants describing the behavior of the tracer to deduce the invariants characteristic of the sytem which is the usual goal of a tracer experiment

The passage of a tracer through a chain of steady state compartments with one-way flow, as in the sys tem $\rightarrow A \rightarrow B \rightarrow C \rightarrow D \rightarrow$, is described with equa tions identical in form to the Bateman equations which describe radioisotopic decay chains (see NUCLEAR REACTION, RADIOACTIVITY) The solution equation for the nth compartment will include n exponential terms, with each exponential constant being equal to one of the n turnover rate constants (the ratio of the rate of entrance or exit of the sub-stance of interest to the amount present in a given compartment) For example, the disappearance of a tracer from the simplest model $\rightarrow A \rightarrow$, assum ing that after time zero no tracer enters is de scribed by

$$x = x_0 e^{-\lambda t} \quad (1)$$

where x is the specific activity in the compartment x_0 is the value at time zero t is time, e is the base of the natural logarithms, 2 718 and λ is the fractional disappearance constant If instead there is no tracer in the compartment at time zero but there is a constant specific activity x_∞ in the inflow from time zero on the build up of activity in the compartment is described by

$$x = x_\infty (1 - e^{-\lambda t}) \quad (2)$$

In both cases λ is equal to the turnover rate constant k Because all of the λs appear in the equation for the last compartment of a chain, all the turnover rate constants may in principle be determined by sampling the last compartment only but this is insufficient for determining the sequence in the chain

When there is flow in both directions, the simple relationship between the ks and λs is lost The differential equations for the 2 compartment, steady-state, closed system $A \rightleftharpoons B$ are

$$dx_A/dt = -k_A(x_A - x_B) \quad (3)$$
$$dx_B/dt = k_B(x_A - x_B) \quad (4)$$

for which, if at time zero $x_A = x_{A_0}$ and $x_B = 0$, the solutions are

$$x_A/x_{A_0} = C_E + C_1 e^{-\lambda t} \quad (5)$$
$$x_B/x_{A_0} = C_E - C_E e^{-\lambda t} \quad (6)$$

where the Cs are constants determined by the amounts of material present C_E is the equilibrium value and C_1 is the difference between the initial value and the equilibrium value of the material present In this example $\lambda = k_A + k_B$ rather than being equal to one or the other turnover rate constant Calculation of the individual turnover rate constants from the experimental data as represented by Eqs (5) and (6) can, however, be achieved through use of Eqs (7)

$$k_{AB} = C_1\lambda \qquad k_{BA} = C_E\lambda \quad (7)$$

In the general case of N compartments in the steady state

$$\frac{dx_J}{dt} = -K_J x_J + \sum_{R=A}^{R=N} k_{JR}x_R \qquad R \neq J \quad (8)$$

describes the behavior of the tracer in the Jth compartment where k_{JR} is the fractional turnover rate constant resulting from exchange with the Rth compartment and K_J is the total turnover rate constant for compartment J The solutions have the general form

$$x_J = C_{J1}e^{-\lambda_1 t} + C_{J2}e^{-\lambda_2 t}, \quad \ldots, C_{JN}e^{-\lambda_N t} + C_E \quad (9)$$

with $N\lambda$s for open systems and $(N-1)\lambda$s for closed systems Interpretation of experimental data in terms of the sizes of and rates of transfer among the various compartments involves fitting Eq (9) to the data and using the constants (the Cs and λs) to calculate the ks and other characteristics of the system Complete sets of formulas for such calculations for the 3 compartment systems are available Because of the unwieldy equations encountered, explicit formulas for deriving the ks from the Cs and λs have not been published for systems involving more than three compartments but numerical examples can be worked with matrix algebra using the general relationship $|k| = |C||-\lambda||C|^{-1}$, where $|k|$ is a matrix of the ks and Ks in Eq (8), $|C|$ is a matrix of the coefficients in Eq (9), $|-\lambda|$ is a diagonal matrix of the $-\lambda$s and $|C|^{-1}$ is the inverse of $|C|$ This relationship is derived by equating the derivatives of Eq (9) to those of Eq (8)

Complicated systems can also be analyzed with electronic analog and digital computers, which eliminate most of the intermediate mathematics involved in translating descriptions of the data into descriptions of the systems Analog computers have the advantage of directness and simplicity of programming, whereas digital computers have the advantage of numerical accuracy

Figure 1 depicts an electrical analog for a 4 compartment steady state system which uses a direct analog between the electrical and biological components The equations which describe the voltage changes in the electrical system are identical in form with those which describe the net tracer movement in the biological system. Adjustment of the electrical parameters therefore provides a direct method for determination of the values of the biological system components For nonsteady state systems, operational amplifiers and an electrical circuit analogous to the differential equation of the biological model may be used to meet the requirement of unequal flow rates in the two opposing directions *See* ANALOG COMPUTER, DIGITAL COMPUTER

Applications. In chemistry the tracer method has been applied to isotopic exchange reactions, chemical kinetics, structural chemistry self-diffusion studies, and analytical chemistry Self-diffusion (really isotopic interdiffusion) has been studied in solids as well as in liquids and gases In activation analysis a sample is subjected to nuclear bombardment in a reactor or accelerator and the induced activity provides an extremely sensitive method for assay of the amounts of certain trace constituents

In biology one of the outstanding achievements in which radioisotopes have played a role has been the use of C^{14} in the elucidation of the metabolic path of carbon in photosynthesis The products produced in the first few seconds following exposure to light have been identified by combinations of paper chromatography and autoradiography Somewhat similarly the extrathyroidal metabolism of iodine the path of iodine in the thyroid gland and other problems of intermediary metabolism have been studied intensively The concept of the dynamic state of cell constituents is largely attributable to discoveries made with isotopic tracers At one time it was thought that concentration gradients across cell membranes depended upon their being impermeable but the use of iso-

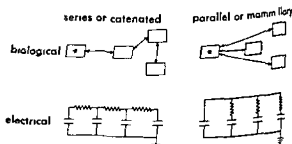


Fig 1 Simple electrical analog scheme for a 4-compartment steady state system The asterisk indicates the injected compartment in the biological system, corresponding to an initial charge on the left hand condenser in the electrical system

es has refuted this hypothesis by proving that many such cases the substances involved are rmally being transported in both directions ross the membrane

In physiology radioisotopes have been used in a ide variety of permeability absorption and distribution studies Representative model systems are epicted in Fig 2 As an illustrative example alcium (Ca) metabolism is described here in some etail Calcium is absorbed from the intestine into he blood from which it is distributed to the soft issues nd skeleton or is excreted via the feces rine i njecte i min reasing 2 months after inje v

Irop in specific activity is explained by dilution of he Ca^{45} in the relatively large pool of rapidly exchangeable bone Ca Although less than 2% of the one Ca is available for exchange this is 10–100 imes the amount in the serum Bone also takes up Ca^{45} by accretion into the nonexchangeable pool In growing boys the rate of bone-salt formation was found to be 2 3 times the rate of Ca intake from the intestine the deficit being made up by resorption The combination of excretion and slow turnover of nonexchangeable bone accounts for the slow components of the serum pecific activity curve In adults the processes of accretion and resorption continue at slower rates and are balanced The distinction between accretion and exchange is possible only with isotopic tracers

Another distinction made possible with Ca^{45} is the finding made by comparing fecal contents following Ca^{45} administration by mouth and by vein that about 10% of fecal calcium is of endogenous origin whereas the rest represents a lack of absorption Comparison of Ca^{45} absorption and excretion in a boy with that in an adult suggests that absorption depends more on intake than on the need for calcium and that calcium balance is regulated by excretion in the urine of calcium absorbed in excess of the body s needs rather than by control of

Tritium is especially use u graphs because the average β particle emitted has a range of only 1μ in tissue giving especially sharp resolution

As illustrated in Fig 3 chromosomes can be labeled with tritium through the use of tritium labeled thymidine a substance which localizes in deoxyribonucleic acid (DNA) which in turn occurs only in the genetic material After the cells were labeled by being grown in the tritiated thymidine medium they were transferred to a nonradioactive medium for growth of the second generation The rate and mechanism of synthesis and degradation of DNA are under intensive study Several lines of study such as the labeling of successive generations of bacteria and the absence of label in the second generation chromosomes marked by the arrows in Fig 3 indicate that DNA is not normally subject to exchange or turnover proc-

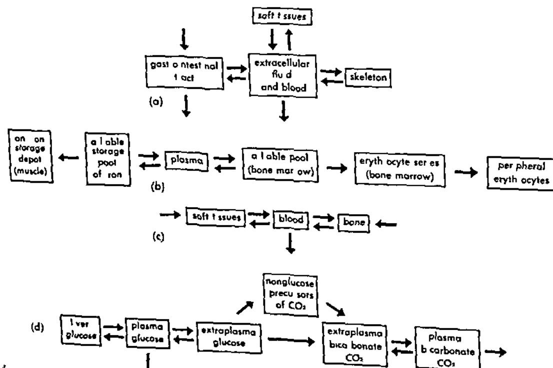


Fig 2 Models used n analyses of metaboliz ng systems (a) Calc um (b) Iron (c) Ca bonate (d) Glucose In (b) ron returns from the per pheral erythrocytes to the plasma very slowly

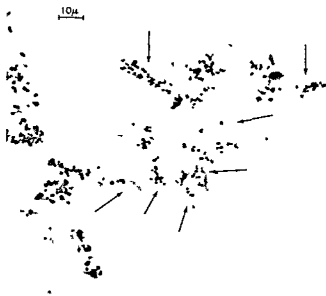


Fig 3 Photomicrograph of an autoradiograph of *Vicia faba* chromosomes in metaphase labeled with tritiated thymidine In the chromosome pairs indicated by the arrows one chromosome is labeled but the other is not This has been interpreted as evidence that the newly formed DNA does not exchange with the previously existing DNA Feulgen stain (*Courtesy W L Hughes*)

esses, but is synthesized only when the cell is pre paring to divide

The kinetics of cellular proliferation has pro vided a rich vein for application of radioisotopic methods For example, the lifetime of human red blood cells (about 120 days) was established with the use of Fe^{59} labeled cells The life cycles of blood cell precursors and of white blood cells and platelets have also been studied with a variety of radioisotopes including tritium

Some applications such as the uptake of I^{131} by the thyroid the measurement of the red cell mass with Cr^{51} labeled red cells and the absorption of Co^{60} labeled vitamin B_{12} are of practical clinical importance in diagnosis and treatment of disease Knowledge of the rates of distribution and disposal of a wide variety of radioactive substances is basic to the problem of evaluation of the hazard from fallout radiation Determination of induced activi ties, particularly Na^{24} provides a basis for calcu lating the neutron exposure in reactor accidents and in therapeutic exposures to neutrons From these few examples it is apparent that aside from their importance in basic chemical and biological investigations, radioactive isotopes will be as im portant in medicine as tracers as they will be as therapeutic agents [J S R O]

Bibliography· C L Comar, *Radioisotopes in Biology and Agriculture* 1955, G Friedlander and J W Kennedy, *Nuclear and Radiochemistry*, 1955, E H Quimby, S Feitelberg, and S Silver, *Radio active Isotopes in Clinical Practice*, 1958, J S Robertson Theory and use of tracer in determin ing transfer rates in biological systems, *Physiol Rev*, 37·133–154, 1957, W E Siri, *Isotopic Tracers and Nuclear Radiations*, 1949, F Stohlman (ed), *The Kinetics of Cellular Proliferation* 1959 A C Wahl and N A Bonner (eds) *Radioactivity Applied to Chemistry*, 1951

Radioisotope production

The preparation for use of radioactive isotopes Because of the usefulness of radioisotopes in medicine, industry, and research much effort has gone into development of techniques for producing and purifying them Major sources of radioisotopes are atoms bombarded with particle-accelerator beams or reactor neutron beams, or the fission products resulting from burned up reactor fuels See RADIOISOTOPE

Some radioisotopes occur in nature for example, uranium, radium, and thorium which were produced when the earth was formed Some of these naturally occurring radioisotopes, or series of radioisotopes, have half lives (time in which one half of the atoms decay) greater than 10^8 years and therefore have not had time to disappear During the early years of work with radioactive materials, only natural radioactivity was available, G Hevesy did the first radioactive tracer experiment in 1934 with radiolead (Pb^{212}) obtained from thorium decay products

rac

α I

activity on a wide scale in scientific work A very wide variety of radioisotopes are produced in particle accelerators such as the cyclotron Charged particles, such as deuterons (D^+) and protons (H^+) are accelerated to great speeds by high voltage electrical fields and allowed to strike targets in which nuclear reactions take place, for example proton in, neutron out (p,n), increasing the target atom atomic number by one without changing the atomic mass, and deuteron in proton out $(d\ p)$ increasing the atomic mass by one without changing the atomic number The target elements become radioactive because the nuclei of the atoms are unbalanced, having an excess or deficit of neutrons or protons Although the particle accelerating machines are most versatile in producing radioisotopes the amounts of radioactive material that can be produced are relatively small that is microcurie to millicurie quantities (a curie is that quantity of a radioisotope required to supply 3.7×10^{10} disintegrations per second)

For large scale production the nuclear reactor with neutron fluxes (amount of neutrons traversing a unit area per unit time) of more than 10^{10} neutrons/(cm²)(sec) is required See PARTICLE ACCELERATION REACTOR NUCLEAR

Reactor production. Radioisotope production steps are shown in Fig 1 Most radioisotopes produced in quantity are made in the nuclear reactor by one of the reactions shown in Table 1

The most common reaction is (1), because many elements have a good neutron capture cross section

Table 1 Nuclear reactions used in radioisotope production

Reaction	Examples
(1) Neutron-gamma $(n\ \gamma)$	$_{27}Co^{59} + {}_0n^1 \to {}_{27}Co^{60} + \gamma$
(2) Neutron proton $(n\ p)$	$_{16}S^{32} + {}_0n^1 \to {}_{15}P^{32} + p^1$
(3) Neutron-alpha $(n\ \alpha)$	$_{17}Cl^{35} + {}_0n^1 \to {}_{15}P^{32} + {}_2\alpha^4$
(4) Neutron fission (n,f)	$_{92}U^{235} + {}_0n^1 \to {}_{52}Te^{131} + {}_{42}Mo^{\sim 103} + \sim 2{}_0n^1$

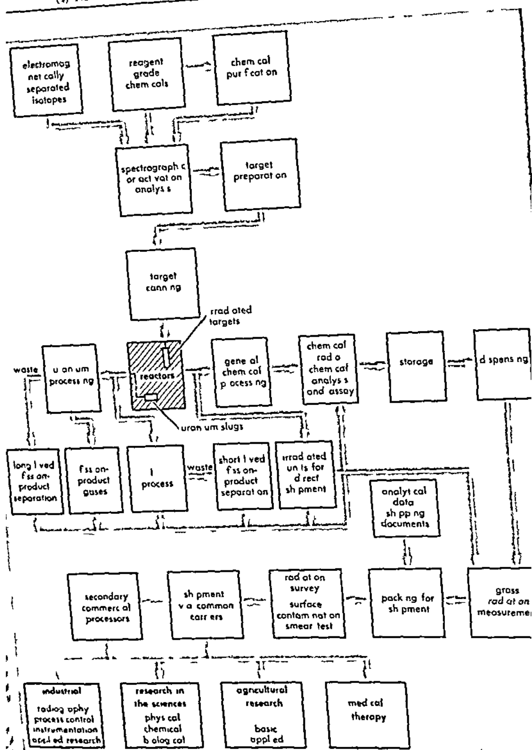


Fig 1 Radio sotope-product on operat ons (Oak R dge Nat onal Laboratory)



Fig 3 Photomicrograph of an autoradiograph of *Vicia faba* chromosomes in metaphase labeled with tritiated thymidine In the chromosome pairs indicated by the arrows, one chromosome is labeled but the other is not This has been interpreted as evidence that the newly formed DNA does not exchange with the previously existing DNA Feulgen stain (*Courtesy W L Hughes*)

esses, but is synthesized only when the cell is preparing to divide

The kinetics of cellular proliferation has provided a rich vein for application of radioisotopic methods For example, the lifetime of human red blood cells (about 120 days) was established with the use of Fe^{59} labeled cells The life cycles of blood cell precursors and of white blood cells and platelets have also been studied with a variety of radioisotopes including tritium

Some applications such as the uptake of I^{131} by the thyroid the measurement of the red cell mass with Cr^{51} labeled red cells and the absorption of Co^{60} labeled vitamin B_{12} are of practical clinical importance in diagnosis and treatment of disease Knowledge of the rates of distribution and disposal of a wide variety of radioactive substances is basic to the problem of evaluation of the hazard from fallout radiation Determination of induced activities particularly Na^{24} provides a basis for calculating the neutron exposure in reactor accidents and in therapeutic exposures to neutrons From these few examples it is apparent that aside from their importance in basic chemical and biological investigations radioactive isotopes will be as important in medicine as tracers as they will be as therapeutic agents [J S RO]

Bibliography C L Comar, *Radioisotopes in Biology and Agriculture*, 1955, G Friedlander and J W Kennedy *Nuclear and Radiochemistry*, 1955, E H Quimby S Feitelberg and S Silver, *Radioactive Isotopes in Clinical Practice*, 1958, J S Robertson, Theory and use of tracer in determining transfer rates in biological systems, *Physiol Rev*, 37 133–154, 1957, W E Siri *Isotopic Tracers and Nuclear Radiations*, 1949, F Stohlman (ed), *The Kinetics of Cellular Proliferation*, 1959 A C Wahl and N A Bonner (eds) *Radioactivity Applied to Chemistry*, 1951

Radioisotope production

The preparation for use of radioactive isotopes. Because of the usefulness of radioisotopes in medicine, industry, and research, much effort has gone into development of techniques for producing and purifying them Major sources of radioisotopes are atoms bombarded with particle-accelerator beams or reactor neutron beams, or the fission products resulting from burned up reactor fuels. *See* RADIOISOTOPE

Some radioisotopes occur in nature, for example uranium, radium, and thorium which were produced when the earth was formed Some of these naturally occurring radioisotopes, or series of radioisotopes, have half lives (time in which one half of the atoms decay) greater than 10^8 years and therefore have not had time to disappear During the early years of work with radioactive materials, only natural radioactivity was available, G Hevesy did the first radioactive tracer experiment in 1934 with radiolead (Pb^{212}) obtained from thorium decay products

Irene Curie produced the first artifically induced radioactivity by irradiating aluminum foil with α particles in 1934, thus initiating the use of radioactivity on a wide scale in scientific work. A very wide variety of radioisotopes are produced in particle accelerators such as the cyclotron Charged particles, such as deuterons (D^+) and protons (H^+), are accelerated to great speeds by high voltage electrical fields and allowed to strike targets in which nuclear reactions take place for example proton in, neutron out $(p\ n)$, increasing the target atom atomic number by one without changing the atomic mass, and deuteron in proton out $(d\ p)$ increasing the atomic mass by one without changing the atomic number The target elements become radioactive because the nuclei of the atoms [illegible] having an excess of

ing radioisotopes the amounts of radioactive material that can be produced are relatively small that is microcurie to millicurie quantities (a curie is that quantity of a radioisotope required to supply $3\ 7 \times 10^{10}$ disintegrations per second)

For large-scale production, the nuclear reactor with neutron fluxes (amount of neutrons traversing a unit area per unit time) of more than $10^{[illegible]}$ neutrons/(cm²) (sec) is required *See* PARTICLE ACCELERATION, REACTOR, NUCLEAR

Reactor production. Radioisotope production steps are shown in Fig 1 Most radioisotopes produced in quantity are made in the nuclear reactor by one of the reactions shown in Table 1

The most common reaction is (1), because many elements have a good neutron capture cross section

Table 1 Nuclear reactions used in radioisotope production

Reaction	Examples
(1) Neutron gamma (n,γ)	${}_{27}Co^{59} + {}_{0}n^{1} \rightarrow {}_{27}Co^{60} + \gamma$
(2) Neutron proton (n,p)	${}_{16}S^{32} + {}_{0}n^{1} \rightarrow {}_{15}P^{32} + {}_{1}p^{1}$
(3) Neutron-alpha (n,α)	${}_{17}Cl^{35} + {}_{0}n^{1} \rightarrow {}_{15}P^{32} + {}_{2}\alpha^{4}$
(4) Neutron fission (n,f)	${}_{92}U^{235} + {}_{0}n^{1} \rightarrow {}_{52}Te^{131} + {}_{42}Mo^{\sim 103} + \sim 2\,{}_{0}n^{1}$

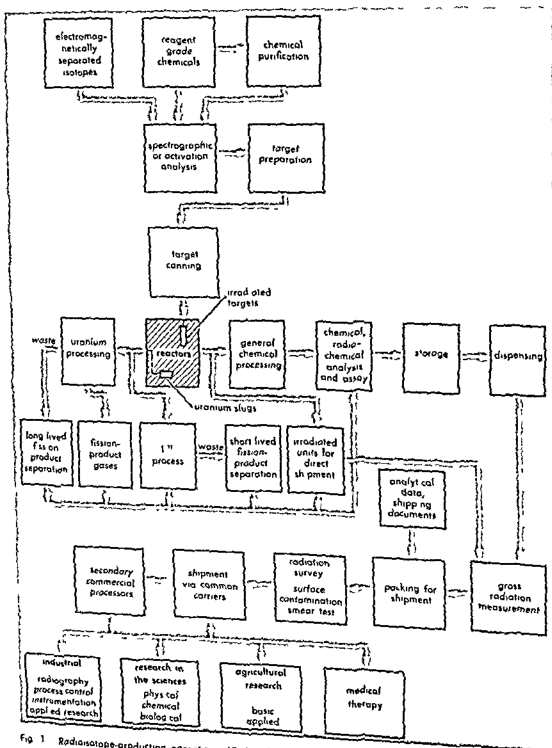


Fig. 1 Radioisotope-production operations (Oak Ridge National Laboratory)

(relative ability for capturing neutrons) By simple neutron capture, important radioisotopes such as Na^{24}, Fe^{59}, Co^{60}, and Au^{198} are made The procedure is quite simple in this case Highly purified materials (to prevent contamination by neutron capturing impurities) in amounts ranging from a few milligrams to several hundred grams are sealed in pure aluminum cans and placed in the reactor In some graphite reactors, the cans are placed in holes in graphite blocks, called stringers which are shoved into the reactor Other reactors have chain conveyors for handling cans and pneumatic or hydraulic tubes for passing samples in and out of the reactor, or special magnesium beryllium, or aluminum holders for insertion into the reactor lattice (Fig 2) Care must be taken to avoid putting materials that decompose easily such as organic compounds, into the reactor because gas pressure may be produced in the container In general materials that are thermally stable are also fairly stable under neutron irradiation Metals or stable oxides of elements are usually used as target materials *See* NUCLEAR REACTION

The production rate for a radioisotope in a reactor depends upon the neutron flux amount of target atom the half life of the radioisotope, and the neutron activation cross section of the target atom The following formula may be used for calculating the amount of radioisotope produced

$$A = N\phi\sigma(1 - e^{-\lambda t})$$

where A is the activity in disintegrations/sec N is the number of target atoms ϕ is the neutron flux in neutrons/(cm²) (sec), σ is the activation cross section for the reaction in cm²/atom and the expression $(1 - e^{-\lambda t})$ is the saturation factor The irradiation time t and the decay constant λ are in compatible units (λ = 0 693 half life) The activity A, in disintegrations/sec, can be converted to millicuries by dividing by $3\ 70 \times 10^7$ disintegrations/(sec) (mc)

Target materials of some elements can be obtained quite pure, when such elements have only one isotope that has a high activation cross section radioisotopes are easy to produce and little chemical purification is required after irradiation In other cases, there is multiple production in the target material sometimes four or five different radioactive species in one target, which must be chemically separated from the main product. Virtually every known chemical separation procedure is used in this kind of work, which usually must be done by remote methods because of the high radiation levels involved *See* NUCLEAR CHEMISTRY

When radioisotopes are produced by (n,p) or (n,α) reactions or an (n,γ) reaction followed by β decay, the radioelement is of a different chemical species from the target element and may therefore be separated from it chemically to produce carrier free radioisotopes

Very often concentrations of radioelements are too low for them to be precipitated directly so they are carried on the surface of a flocculent precipitate, such as $Fe(OH)_3$, similarly others are coprecipitated where isomorphous compounds are brought down together for example, $Ba^{140}SO_4$ with $PbSO_4$ Many newer methods, more adap-

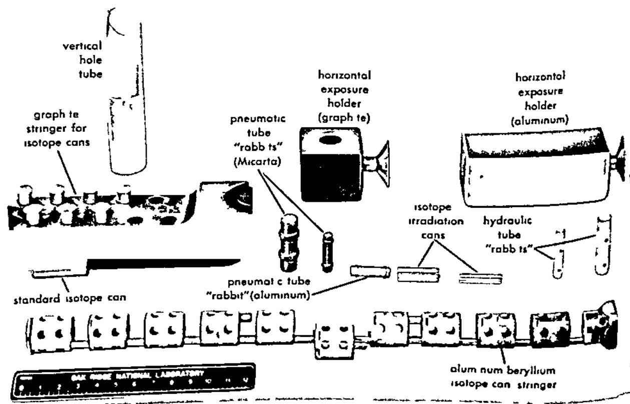


Fig 2 Devices for holding samples for irradiation in a reactor (*Oak Ridge National Laboratory*)

ble to work with low concentrations of material or amenable to remote operation are now used such as solvent extraction and ion exchange A preferred purification method is gasification because if the radioisotope goes through various stages in the gaseous phase very high purity usually results Distillation is used in some cases such as the purification of I^{131} (distilled as the element) and — ^{106} (distilled as the tetroxide) Ion ex — ...tically the only method for fraction

...

fragments of the fissioned uranium or plu... m atom are an extremely important group of isotopes ranging from zinc (atomic no 30) amarium (atomic no 62) The fission product ments occur in two groups, one being light ns with atomic masses between 72 and 110 s and the other being heavy atoms with masses ween 110 and 162 units Those fission products h high yields occur at one of the peaks, and the apanion fission fragment occurs with the same h yield on the other peak for example mass) (Xe through Ba) has companion fragments of out mass 95 (Y through Mo) Fission product elds are given as the per cent of fissions that sult in fragments of a certain mass there are o fragments for each fission so the over all eld arithmetically totals up to 200% See Fis- ON NUCLEAR

The fission products ordinarily separated and urified for use as radioisotopes fall into two groups, as listed in Table 2

The short lived fission products have been sepa rated principally for research purposes Several methods have been used for example inclusion of carrier elements in classical precipitation methods

Uranium metal freshly discharged from the re- actor is dissolved in 70% nitric acid During dis- solution the volatile materials such as iodine and

Table 2. Isotopes ordinarily separated from fission products

Radioisotope	Half life	Fission yield %
Principal long lived fission products*		
Kr^{85}	10 27	1 3
Sr^{90}	28	5 8
Tc^{99}	212 000	6 1
Ru^{106}	1	0 4
Cs^{137}	30	5 9
Ce^{144}	0 8	5 7
Pm^{147}	2 6	2 4
Principal short lived fission products†		
Sr^{89}	54	4 8
Y^{91}	59 5	5 8
Zr^{95}-Nb^{95}	65–35	6 3
Ru^{103}	39 8	3 0
I^{131}	8 08	[illegible] 9
Xe^{133}	5 3	6 5
Ba^{140}	12 8	6 4
Ce^{141}	32 5	6 0
Pr^{143}	13 9	6 0
Nd^{147}	11 1	2 4

* Half life in years † Half life in days

... Kr and Xe are released Iodine is ...

rare gases are purified by absorption ... taminating nitrogen oxides in the gas stream, and the Kr and Xe are then absorbed at low tempera ture on an activated charcoal column The noble gases are eluted from the column by heating and passing a stream of helium through the column

The nonvolatile fission products remain in the uranium nitrate solution Uranium can be removed by precipitation or ion exchange However a better method is solvent extraction with a mixture of kerosine and tributylphosphate (TBP), many other solvents which dissolve uranium selectively can also be used. Nitric acid (about 4 M) can be used as a salting-out agent with TBP in other cases alu minum nitrate solution is used

Zr^{95}-Nb^{95} exhibits colloidal properties in the dissolver solution and can be removed directly and selectively by passing the uranium fission product solution through a column packed with silica gel The Zr^{95}-Nb^{95} is then eluted with oxalic acid solution

After uranium is extracted the rare earths Ce^{144} Nd^{147} Pr^{143} and Y^{91} and alkaline earths Sr^{89} and Ba^{140} are separated from the solution by ion exchange Ru^{103} is removed by distillation as the tetroxide

The starting material for production of long lived fission products is the waste from reactor fuel processing plants In the ideal case the fis sion product mixture is in a fairly pure nitric acid solution which may be evaporated to allow greater plant through put However it must not be evapo- rated beyond the point where the total radioactivity per unit volume (curies/ml) is too high for sub- sequent processing because of radiation and radio- active heating problems With concentrations as high as several curies per milliliter it is possible to precipitate the fission products directly

Cs^{137} is separated this way at the British and ... precipita ... cesium is ... American processes for cesium include precipitation as a double salt of nickel cobalt or zinc ferrocyanide or cocrystallization with ammonium alum The lat ter process is highly selective and allows removal of cesium from complex chemical mixtures

In most of the known flow sheets for large scale fission product separation the alkaline earth and rare earth groups are first precipitated together in an alkaline carbonate solution The iron and other contaminating corrosion products are usually re- moved prior to this step by a selective hydroxide precipitation with ammonium hydroxide by care fully controlling the pH Sr^{90} is included in the alkaline carbonate precipitate along with fairly large amounts of stable barium and strontium and rare earths The rare earths are precipitated as hydroxides from a carbonate-free solution allow ing the alkaline earths (Sr and Ba) to pass into the filtrate Strontium and barium can then be pre

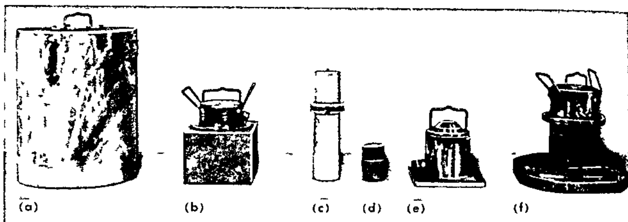


Fig 3 Radioisotope shipping containers (*a*) Lead shield, stainless-steel encased for gamma sources, such as Co^{60} (*b*) Lead shield, stainless steel encased, in wooden block for pile units and small sources (*c*) Gas shipping container with internal sealed cylinder for Kr^{85} and tritium gas (*d*) Nonreturnable lead shield for processed radioisotope bottles (*e*) Solid stainless-steel container for bottles of processed Sr^{90} (*f*) Lead shield, stainless steel encased, on a shipping pallet for distributing weight in aircraft, for pile units and small gamma sources (*Oak Ridge National Laboratory*)

cipitated as the carbonates and purified by reprecipitation from 80% HNO_3 For production of Sr^{90} beta sources, it is necessary to remove the barium to get higher specific activity This is done by converting the Sr Ba to chlorides in strong HCl solution and selectively precipitating $BaCl_2$ by raising the HCl concentration to 12–13 *M* Strontium can then be precipitated as SrF_2 or $SrCO_3$ Similar methods can be used on the other fission products



Fig 4 Nonreturnable shipping container for radioisotopes (*Oak Ridge National Laboratory*)

Precipitation processes are also sometimes supplemented or replaced by very selective ion-exchange or solvent extraction methods

Radioisotope shipping. The packaging and shipment of radioisotopes present some unusual problems—the penetrating radiation must be shielded and leakage must be prevented because of the extreme toxicity of many radioisotopes

Packaging must be done by remote control Most radioisotopes are dispensed as water solutions by pipetting from glass storage bottles to glass shipping bottles [illegible] used with clos[illegible] For some special preparations, suc[illegible] Ca^{45} or P^{32}, polyethylene storage and shipping bottles are used to cut down losses by adsorption on the walls of the container High standards of cleanliness must be maintained, although no attempt is made to ensure sterility of solutions except for the special medical radioisotope preparations shipped by pharmaceutical suppliers

A few preparations have sufficient mass per unit of radioactivity and weak enough radioactivity to permit handling and packaging them as dry solids. An example of this is C^{14} which is dispensed as barium carbonate Materials that are not chemically processed such as reactor irradiation units or metallic cobalt are sent out as solids in metal containers

Two general kinds of shipping container are used returnable on which a deposit is paid and nonreturnable which the receiver may keep and use Returnable containers are usually more massive and expensive especially those used for shipping large amounts of Co^{60} or fission products. Pure β emitters such as P^{32} require little shielding other than the absorbent packing material. Gases are shipped in special glass or metal containers (Figs 3 and 4)

Shipment of radioactive materials by rail or truck is covered by ICC regulations, air shipments are controlled by the CAA and water shipments

the Coast Guard Detailed regulations are published by the U S Bureau of Explosives See ISOTOPE SEPARATION (STABLE ISOTOPES), RADIOACTIV Y RADIOISOTOPE (ASSAY), TRACER, RADIOACTIVE [A F RU]

Bibliography Atomic Energy Commission *Handbook of Federal Regulations Applying to Transportation of Radioactive Materials* 1958, H Etherington (ed), *Nuclear Engineering Handbook* 1958, *Proc Intern Conf Peaceful Uses Atomic Energy*, vol 14 1956, *Proc Second Intern Conf Peaceful Uses Atomic Energy*, vol 4 1958

Radiolarian earth

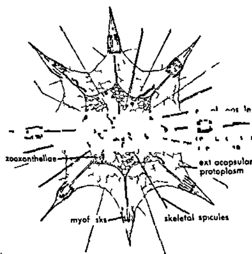
A por
from
Formed from rad

e and cream colored Minor components in rian earths are manganese nodules shark and other resistant remains of vertebrates. arites are known from Devonian to Terti e *See* CHERT, RADIOLARIDA, SEDIMENTARY [R S]

olarida

rder of marine protozoa in the class Actino-

pene
tainin
larida
l to m
side the central capsule Radiolarida are sometimes enclosed in a gelatinous sheath penetrated by



Acanthometra one of the Act pyl na (after Moroff and St asny) (From L H Hyman The Invertebrates vol 1 McGraw H ll 1940)

posits of rad
have been formed from such deposits, rad
skeletons are of value in geological correlation [R P H]
See ACTINOPODEA

Radiology

The medical science concerned with radioactive substances, x rays, and other ionizing radiations, and the application of the principles of this science to diagnosis and treatment of disease

Diagnostic radiology. This technology is based largely on the use of radiations, which are mainly x rays, to study the anatomical configuration or function the use of radioisotopes
cluded diagnostic radiology, but the following are the most common

Radiography is the making of shadow images on photographic emulsion by the action of ionizing radiation The image is the result of the differential absorption of the radiation in its passage through the part of the body being radiographed *See* RADIOGRAPHY

Fluoroscopy is the examination of the internal structure of a part of the body by passing a beam of ionizing radiation, which is usually x rays through
rescent
permits
In some cases such as gastrointestinal examination contrast can be increased by administering a dense fluid such as a suspension of barium sulfate, to the patient Other contrast media are used gall

trast) gives very good contrast but is dangerous because it remains in the body for many years and its nuclear radiations may eventually cause tissue damage and cancer in the organs in which it remains or concentrates The danger in this case arises from the radioactivity of thorium and its daughter products although this material is used because it has a high atomic number and not because it is radioactive Other contrast media may cause severe reactions in some patients for other reasons

Photofluorography consists in photographing the visual image on the fluorescent screen usually with much reduction in size There are many other techniques in diagnostic radiology with special names indicating usually the organs under examination such as angiocardiography (heart and great vessels) cholecystography (gall bladder), and others

Image intensification or amplification is brought about by special electronic devices which permit visualization of internal organs with x ray beams of much lower intensity than are required in direct

fluoroscopy This greatly reduces the dose received by the patient during an examination The technique has other practical advantages, mainly of convenience

The best example of the use of a radioisotope for diagnostic purposes is the study of thyroid function by determining the uptake of radioactive iodine (usually I^{131}) by the thyroid gland after the material has been administered orally to the patient The accumulation of iodine in the thyroid may be determined by a suitable γ ray detector, such as a Geiger Muller counter, placed near the gland More elaborate devices, such as electronic scanners are used to determine the distribution of γ ray emitting isotopes in certain body regions or organs

Therapeutic radiology Radiation therapy deals with the treatment of disease with ionizing radiation The diseases most commonly treated are cancer and allied diseases Radiation therapy has been found useful in the management of some diseases such as ringworm of the scalp and bursitis but because of possible serious complications occurring many years later the use of ionizing radiation is generally avoided if alternative methods of treatment are available

In cancer therapy the objective is to destroy a tumor without causing irreparable radiation damage in normal body tissues that must of necessity be irradiated in the process of delivering a lethal dose to the tumor This applies particularly to important normal structures in the vicinity of the tumor The relative radiosensitivity of the tumor with respect to these normal structures is the chief factor determining the success of the treatment The effect of the radiation on the tumor and other tissues becomes apparent gradually, in a matter of weeks During this latent period a certain amount of recovery or repair may also take place

The optimal differential between the effect on the tumor and the effect on normal tissues or the patient as a whole results from the proper adjustment of many treatment factors which in general requires great clinical experience The time factor is very important This involves the administration of the therapeutic dose in one of three ways (1) by one short treatment (2) by protraction as continuous irradiation over a long time or (3) by fractionation in small repeated doses The importance of the distribution of radiation in the patient s body is obvious Ideally only the tumor should be irradiated but this is impossible for one thing because in general there is infiltration of adjacent normal tissues of unknown extent If the radiosensitivity of the tumor is much higher than that of the surrounding normal tissue the problem is relatively simple but in general this is not the case

Various techniques have been devised to bring about a reasonable distribution of radiation in the region of the tumor When x rays are used for the treatment of deep seated tumors the beam must traverse a considerable thickness of normal tissues To minimize absorption in these tissues, higher and higher voltage x rays have been used Supervoltage x ray therapy makes use of x rays produced at several hundred kilovolts (kv) to 2000 kv Megavoltage x rays range from a few to 70 megavolts (Mv) Cross fire treatment involves the use of more than one beam aimed at the tumor, but passing through different skin areas Rotation therapy involves the use of one beam constantly aimed at the tumor while either the patient or the source of radiation is rotated about an axis passing through the center of the tumor When the source is rotated, 100-2000 curies of cobalt 60 (Co^{60}) are generally used to provide a well collimated beam of γ rays (see CURIE) Fast neutrons, high energy electrons and deuterons as external beams have been used for therapeutic purposes Using radioactive substances other techniques are possible In the case of intracavitary therapy, the source, in a suitable container, is placed within a body cavity in which cancer is present In the case of interstitial therapy many needlelike sources are distributed throughout the tumor and removed at the completion of the treatment Using radon which has a short half life the sources may be made into radon seeds which are very small and may be left in place permanently Sources of tantalum 182 (Ta^{182}) are sometimes used in a similar way Attempts to infiltrate a tumor with radioactive material in fluid form have not been very successful but some good results have been obtained by use of colloidal gold 198 in the treatment of cancer of the prostate The same material has been used for the palliative treatment of ascites tumors in the pleural or abdominal cavity Phosphorus 32 has been found useful in the treatment of polycythemia vera and of leukemia by injection into the blood stream Iodine 131 which when introduced into the body by any route tends to concentrate almost entirely in the thyroid is very effective in reducing the size and function of the thyroid in patients suffering from [illegible] of cancer of [illegible]

[illegible] larger and the arrest of the tumor process is generally temporary

Radiation therapy involving the production of ionizing radiation within the tumor itself offers some possibilities but has not been explored sufficiently The method requires first the localization in the tumor of a substance that can be made to emit ionizing radiation by exposure to neutrons Boron 10 and lithium 6 are two such substances Uranium might also be used but the problem of localization is more difficult [G.F.]

Radiometer, acoustic

[illegible]

the Rayleigh disk This consists of a thin disk of

agnitude

$$r = \frac{4}{3}\rho r^3 \xi^2$$

where ρ is the density of the medium and ξ the particle velocity This formula holds for wavelengths that are large compared to the diameter of the disk The formula has been quantitatively satisfied under many conditions so that the Rayleigh disk offers a well tried and useful method for the calibration of sound fields in air

Because of the high density of water which is comparable with the density of the disk the formula needs correction when used to measure sound intensity in water A B Wood has derived the formula

$$r = \frac{4}{3}\rho_w r^3 \xi^2 \left(\frac{m_e - m_w}{m_e + m'}\right)^2$$

in which m_e is the mass of the disk, m_w the mass of the water displaced by the disk and $m' =$ 3) $\rho_w r^3$ where ρ_w is the density of water

Other radiometers use the pressure of sound waves to deflect a spherical body Since the radiation pressure is p 2E where E is the energy of the acoustic wave per unit area a direct measurement of this pressure will determine the energy of plane wave One of the most common types of acoustic radiometer is the torsion vane pendulum, which gives an angle of displacement proportional to the radiation pressure Although sound radiometers are rather insensitive, they provide a convenient method for measuring the intensity of continuous sounds of large amplitude [W P M]

Radiometric assay

An analytical technique that includes procedures for measuring by tracer methods elements which are not themselves radioactive It is particularly useful for small quantities of inorganic ions or complex organic compounds A typical procedure giving sensitivities comparable to spectroscopy involves quantitative paper chromatography of traces of metals and subsequent exposure to hydrogen sulfide labeled with sulfur-35 Measurement of radiations from the highly insoluble radioactive sulfides is more sensitive than color tests Complex organic molecules are often treated with iodine-131 labeled pipsyl chloride in this same manner

In radiometric titrations a radioactive tracer is used to determine the equivalence point in a volumetric determination where a highly insoluble compound of definite composition is formed At least one of the solutions used in the titration must be radioactive because progress of the titration is followed by plotting the radioactivity of the solution versus volume of reagent added The equivalence point is determined by the intersection of two straight lines on this plot because a definite change in solution activity appears after equivalent amounts of reagents have reacted Submicrocurie levels of radioactive tracer suffice for most applications See RADIOISOTOPE (ASSAY), TITRATION, TRACER RADIOACTIVE. [W ME.]

Radiometry

The detection and measurement of radiant electromagnetic energy Conventionally, radiometry is concerned with infrared radiation Generally, the devices used in radiometry can be and are, used with visible light However the use of devices applicable only to visible light is commonly excluded from the term radiometry For a summary of the methods used for detection and measurement of the electromagnetic spectrum as a whole see ELECTROMAGNETIC RADIATION See also INFRARED RADIATION, PHOTOMETRY

In 1800 Sir William Herschel studied the ability of sunlight to heat a sensitive mercury in glass thermometer He detected radiation beyond the red end of the visible spectrum hence the name infrared He also noted that radiation could be detected from moderately hot objects showing no visible light Detectors such as the thermometer which respond to the increase in temperature resulting from the absorption of radiant energy are termed t

fect does not depend upon a rise in temperature, but upon the freeing of a bound electron by the absorption of a single quantum of radiation Detectors utilizing this principle are termed quantum detectors or photodetectors

Thermal detectors. The mercury in glass thermometer is sluggish and relatively insensitive So also is the Crookes radiometer, which consists of two blackened vanes at the ends of a horizontal rod suspended from a fine quartz fiber in an atmosphere of air or other gas at about 0 1 mm Hg pressure Radiation absorbed by one of the vanes heats it, and gas molecules recoiling with added momentum from the warm vane, tend to turn it about the suspension The Crookes radiometer survives in jewelers' windows as a "perpetual motion" device

Improvement in thermal detectors has been concerned with securing a large and rapid rise in temperature and high sensitivity in the detection of changes in temperature. The temperature of any thermal detector will increase until the rate of loss of heat to its surroundings is equal to the rate at which radiant energy is absorbed To secure the largest rise for a given radiation, the detector should absorb it as completely as possible, the loss of heat, by reradiation to its surroundings, by conduction through its supports, and by gas conduction and convection, must be as small as possible. The heat capacity of a thermal detector should be small so that a rapid rise in temperature will occur The area of the absorbing surface should be small to permit the measurement of narrow beams

of radiation These criteria lead to detectors which are very thin, with thin black coatings, supported commonly by two wires Enclosing the detector in a vacuum decreases the heat loss, but requires a window capable of transmitting the radiation

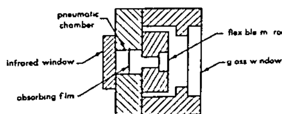
The thermocouple, one of the oldest and still one of the best radiation detectors, produces an electro motive force (emf) when heated The sensitive por tion of a bolometer undergoes a change in resist ance when heated For a discussion of these de vices, *see* BOLOMETER, THERMOCOUPLE

The essential part of a Golay pneumatic radiom eter, a type of thermal detector, is shown in the figure An air or gas filled cell has at its center a radiation absorbing film The gas, heated by the ab sorption of radiation, expands and changes the curvature of the mirror This mirror consists of a thin plastic film coated by evaporation with a layer of antimony A beam of light reflected from this mirror to a photocell permits slight changes of curvature in the mirror to be measured by means of changes in reflected light The Golay cell is in tended for use in a chopped beam of radiation (To secure the alternating potential desired for amplification, it is general practice to "chop" the radiation with a rotating sector disk or by some other means)

Quantum detectors. Photoemissive cells and photomultipliers are presently of interest for visi ble radiation, but do have limited application in the near infrared to about 1 μ (*see* PHOTOTUBE, PHOTOTUBE, MULTIPLIER) Photoconductive cells using materials such as lead sulfide, lead selenide, and lead telluride are useful to 5 μ Photoconduc tive cells of single crystals of gold doped germa nium, cooled with liquid helium, are sensitive out to much longer wavelengths (*see* PHOTOCON DUCTIVE CELL, PHOTODIODE, PHOTOTRANSISTOR)

Such quantum detectors have short time con stants, permitting high chopping rates, and are su perior to thermal detectors for following rapidly changing radiation flux Their sensitivity is more dependent on wavelength, and they are not oper able as far into the infrared as are thermal detec tors

Performance of detectors. Interest in detection and measurement of radiation arises in a great va riety of circumstances for which a variety of detec tors are suitable or adequate For full sunlight, a receiver of large area, consisting of a number of thermocouples in series and known as a thermo



Schematic diagram of the detecting part of a Golay pneumatic radiometer

Characteristics of thermal detectors

Type	Thermo-couple	Bolometer	Golay pneumatic
Material	Bi Sb vs Bi Sn	Platinum	Gas-filled
Time constant, sec	0 036	0 016	0 015
Area mm^2	*0 5*	*1 6*	*8 0*
Frequency of measurement, cps	0 5	10	10
Resistance, ohms	5	40	
Noise-equivalent power, watts	$0\ 5 \times 10^{-10}$	$1\ 7 \times 10^{-10}$	$1\ 5 \times 10^{-10}$

pile, together with a simple galvanometer may be adequate For the radiation of a star, a single ther mocouple and an amplifier might be used. For rapid scanning of an infrared molecular spectrum, a maximum of sensitivity and speed of response are desirable

The response of thermal detectors is generally independent of wavelength over a range from the ultraviolet to wavelengths of the order of magni tude of the dimensions of the receiver Quantum detectors generally have maximum sensitivity in the visible or near infrared regions, and are unre sponsive above a certain cutoff wavelength

It might appear that amplification could be in creased as much as desired by the use of an elec tronic amplifier However, as amplification of any signal by any type of amplifier is increased random fluctuations in the signal (noise) also will be in creasingly amplified For radiation detection the limiting noise is not in the amplifier but in the de tector itself Thermocouples and bolometers are generally limited by so called Johnson noise (also called thermal noise) which is caused by thermal agitation of electrons and which has a magnitude given by

$$e^2 = 4kTR(f_1 - f_2)$$

where e = root mean square (rms) noise voltage
k = Boltzmann constant
T = absolute temperature
R = electrical resistance of detector
$(f_1 - f_2)$ = frequency limits between which rms volt age is measured

See NOISE ELECTRICAL

A commonly used measure of the sensitivity of a detector is the noise equivalent power that is, the watts of radiation which produce a response equal to the noise for an amplifier bandwith $(f_1 - f_2)$ of 1 cps This and some other characteristics of three kinds of thermal detectors are given in the accom panying table [H W R.]

Bibliography W E Forsythe (ed) *Measurement of Radiant Energy* 1937 D E Gray (ed) *American Institute of Physics Handbook* 1957 R A Smith F E Jones and R P Chasmar *The Detection and Measurement of Infra red Radiation* 1957

micrometer

strument consisting of a combination of a ocouple and a galvanometer used for meas quantity of infrared radiation The radio- ented by C V Boys in 1887 to

BOLOMETER GALVANOMETER
RMOCOUPLE [H W R]

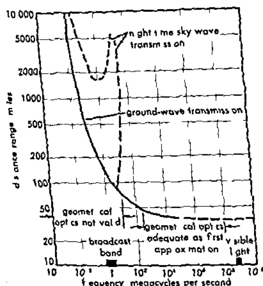
lio wave propagation

es which travel with the velocity of ... be reflected refracted diffracted scattered, and

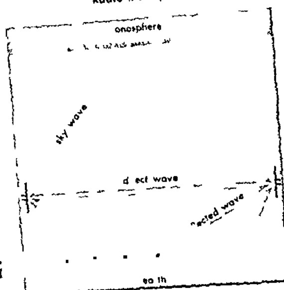
ency is an important parameter radio propaga on characteristics vary over a wide range At the igher radio frequencies the similarity with vis le light is evident At the lower frequencies the adio waves follow the surface of the earth by a nechanism that in geometrical optics is unimpor tant and relatively unknown

The power radiated from a transmitting antenna is ordinarily spread over a relatively wide area As a result the power available at most receiving an tennas is only a very small fraction (10^{-8} 10^{-16} or

der to predict the usefulness of a radio signal at any location remote from the transmitter For a dis cussion of antenna radiation *see* ANTENNA (AE RIAL) *see also* ELECTROMAGNETIC RADIATION



Fg 1 Transm ss on range of electromagnet c waves



Fg 2 Poss ble transmiss on paths between two an tennas

Typical distances that can be achieved with usual types of equipment are shown in Fig 1 The fre quencies around 20 kc can be received reliably at distances of thousands of miles but are limited to telegraph type signals and require very large trans- mitting antennas Higher frequencies are needed for voice and still higher frequencies are needed for television transmission As the frequency in

distances with the usual type of equipment How ever distances of 200 miles or more are possible by the use of high power and large antennas to pro- vide narrow searchlight beams

Reflections from the ionosphere (ionized layers 50–200 miles above the earth s surface) provide a useful but variable long distance service at fre- quencies less than about 30 Mc These reflections account for the long range broadcast coverage at night and for the short wave intercontinental com munication *See* IONOSPHERE

The principal components of the received radio signal are shown symbolically in Fig 2 The vector sum of the direct, reflected and surface waves has been called the space wave ground wave or tropo- *spheric* transmission to differentiate it from the ion ospheric reflections The ionospheric and surface wave are the principal *components* at frequencies below 10–30 Mc The direct and reflected rays are the principal factors at frequencies above 30–50 Mc While the ionospheric direct and ground reflected waves can be easily visualized the surface wave is more difficult to understand it is a *correction* fac tor' that arises because the earth is *not* a perfect reflector

Line of sight transmission This is the first and simplest *concept* in radio-wave propagation Radio *transmission* in free space results in a decrease in

energy per unit area in accordance with the inverse square law (*see* INVERSE SQUARE LAW) The ratio of the transmitter power to the received power is called the radio transmission loss, and for nondirectional antennas in free space it is $P_T/P_0 = (4\pi d/\lambda)^2$ where the distance d and wavelength λ are expressed in the same units Here P_T is the transmitted power and P_0 is the power received under free space conditions

Ground effects The presence of the ground affects the generation and propagation of radio waves so that the received power is ordinarily less than would be expected in free space At low frequencies (or when one or both antennas are connected to earth), the received power P_R is less than the free space value by the fraction

$$\frac{P_R}{P_0} = \left| K\frac{\lambda}{d} \right|^2 \qquad (1)$$

The factor K is greater for vertical polarization of the waves than for horizontal polarization and is greater for transmission over sea water than for transmission over dry land

At frequencies in the television band and higher both the transmitting antenna and the receiving antenna are usually several wavelengths above the ground and the power P_R received over flat earth is given by

$$\frac{P_R}{P_0} = \left| 2 \sin \frac{2\pi h_1 h_2}{\lambda d} \right|^2 \qquad (2)$$

where h_1 and h_2 are the antenna heights in the same units as the wavelength and distance

At still higher frequencies that is in the microwave region the irregularities in the surface of the earth are frequently large compared to the wavelength and the magnitude of the reflection coefficient becomes substantially less than unity (*see* REFLECTION AND TRANSMISSION COEFFICIENTS) As the reflected wave is weakened radio propagation approaches free space transmission

Fading Variations in signal level with time are caused by changing atmospheric conditions The severity of the fading usually increases as either the frequency or path length increases The path of a radio wave is not a straight line except for the ideal case of a uniform atmosphere The transmission path may be refracted up or down depending on atmospheric conditions This bending may either increase or decrease the effective path clearance and unfavorable (inverse) bending may have the effect of transforming a line of sight path into an obstructed one This type of fading may last for several hours at a time The frequency of its occurrence and its depth can be reduced by increasing the path clearance, particularly in the middle of the path

Most of the fading that occurs on "rough" paths with adequate clearance is the result of interference between two or more rays traveling slightly different routes in the atmosphere This fast multipath type of fading is relatively independent of path clearance

Most fading is a temporary diversion of energy to some direction other than the expected location but absorption effects are important in the microwave region At frequencies above 5000–10 000 Mc the presence of rain snow, or fog introduces an absorption in the atmosphere that depends on the amount of moisture and on the frequency At frequencies higher than 15 000–20 000 Mc the additional attenuation caused by heavy rain tend to limit the path length to only a few miles, if high reliability is required In addition to the effect of rain some selective absorption may result from the oxygen and water vapor in the atmosphere. The first absorption peak due to water vapor occurs at about 24 000 Mc and the first absorption peak for oxygen occurs at about 60 000 Mc *See* ABSORPTION (ELECTROMAGNETIC RADIATION)

Beyond horizon transmission This type can be achieved by three principal methods refraction diffraction and reflection When these effects are intermixed and cannot be separated easily the energy is said to be scattered

Refraction The dielectric constant of the atmosphere normally decreases gradually with increasing altitude The result is that on the average the radio energy is bent or refracted toward the earth so that the radio horizon is slightly greater than the optical horizon Since the amount of refraction is variable exceptionally long range transmission may occur occasionally The corresponding phenomenon in optics makes visible lights or other objects that are normally below the horizon Conversely when the radio energy is bent away from the earth (inverse bending) the transmission loss is increased *See* REFRACTION OF WAVES

Diffraction Radio waves are also transmitted around the earth by the phenomenon of diffraction. Diffraction is a fundamental property of wave motion which indicates that the transition from light to dark at the edge of a shadow is gradual rather than infinitely sharp The amount of energy diffracted around an obstruction decreases as the frequency is increased Typical obstructions include hills trees and buildings as well as the curvature of the earth *See* DIFFRACTION

Most of the experimental data at points far beyond the horizon are intermediate between the values expected for diffraction over a smooth sphere and for diffraction over a ridgelike obstruction. Various theories have been advanced to explain these effects but none have been reduced to a simple form for everyday use The explanation most commonly accepted is that energy is reflected or scattered from turbulent air masses At points far beyond the horizon the long term median signal level decreases as the first power of the frequency and as the seventh or eighth power of distance. While useful signals can be obtained at distances of 200 miles or more at all frequencies up to the 5000–10 000 Mc limit set by rain attenuation the

optimum frequency range is below 1000 Mc for most applications The so called beyond horizon or tropospheric scatter circuits require very high power and large antennas but are economically feasible in isolated areas and for over water paths Rapid fading occurs nearly all of the time but with ample [illegible] margin and the use of diversity reception [illegible]

[illegible] to provide an acceptable grade of monochrome tele[illegible]n on a single link of about 150–200 miles

[illegible]flection Ionospheric reflections return to earth [illegible]l radio energy at frequencies up to 25 Mc possibly up to 60 Mc Information about the [illegible]re of the ionosphere has been obtained by [illegible]smitting radar type signals directly overhead by recording the intensity and time delay of the [illegible]es returned from the ionized layers At night frequencies below the critical frequency of 2–4 (for vertical incidence) are returned to earth [illegible]h a received power that is close to the value [illegible]t would be expected in free space for the round [illegible] distance During the day the critical fre[illegible]ncy is 2 3 times greater than the corresponding [illegible]ht value This apparent increase in useful [illegible]quency range is largely offset by the strong day [illegible]e ionospheric absorption which reaches a maxi[illegible]m in the 1 2 Mc range at about noon The differ[illegible]ce between day and night transmission means that [illegible]ost sky wave circuits require two or more fre[illegible]encies for reliable service at all hours

For oblique incidence the maximum usable fre[illegible]uency (MUF) is greater but even at the longer [illegible]istances it does not exceed 3–3.5 times the criti[illegible]al frequency for vertical incidence at that time [illegible]eason and latitude

The frequencies most suitable for transmission over distances of 1000 miles or more will ordinarily not be reflected at the high angles needed for much shorter distances As a result the range of sky wave transmission usually does not overlap the range of ground wave transmission and the intermediate region of very weak or undetectable signals is called the skip zone At frequencies around 1 2 Mc the ground wave and sky wave may overlap with the result that severe fading occurs in the region where the two signals are comparable in amplitude

In addition to the diurnal variations there are systematic changes with season latitude and the 11 year sunspot cycle For example on a summer night in middle latitudes during a year of low sun spot activity the MUF for the longest distance may be limited to less than 7–8 Mc On the other hand during a year of high sunspot activity the corresponding MUF for a winter afternoon may be 40 Mc or higher

In addition to the normal daytime absorption there is a second type of absorption which can occur either day or night and which is particularly troublesome on transmission paths that travel near or through the polar regions During periods of magnetic storms, the auroral zones expand over an [illegible] larger than normal and thereby disrupt [illegible]

[illegible]

In addition to the regular ionospheric reflections, strong signals sometimes occur at frequencies as high as 60–70 Mc from the E layer, which is a region of relatively high ionization located about 50–70 miles above the earth's surface These reflections are known as Sporadic E signals because they are erratic in both time and space

All ionospheric circuits are subject to rapid multipath fading with echo delays up to several milliseconds These delays are about 10 000 times as long as for tropospheric transmission As a result of these relatively long delays, uncorrelated selective fading can occur within a few hundred cycles, this produces the well known distortion on voice circuits that is characteristic of short wave transmission

Scatter transmission As the frequency is increased above that normally reflected from the ionosphere the signal intensity decreases rapidly, but it does not drop out completely Although the signal intensity is low reliable transmission can be obtained at frequencies up to 50 Mc or higher and to distances up to at least 1200–1500 miles Such ionospheric scatter circuits require much higher power and larger antennas than are ordinarily used in ionospheric transmission Ionospheric scatter transmission is suitable for continuous transmission of a few telegraph channels or for one voice circuit but the useful bandwidth is limited by the severe selective fading that is characteristic of all ionospheric transmission Ionospheric scatter is apparently the result of reflections from many patches of ionization in the E region

Momentary peaks of the ionospheric scatter signal intensity occur every few seconds and seem to be the result of the relatively strong ionization produced by the passage and disintegration of small meteors Meteor burst communication transmits information only during the peaks and is silent when the signal intensity is significantly less than its maximum value The required transmitter power is much less than for ionospheric scatter systems but the meteor burst method is limited to discontinuous telegraph type messages of relatively low information capacity

Noise. The usefulness of a radio signal is limited by noise The noise may be either unwanted external interference or noise originating in the receiver itself Atmospheric static is caused by lightning or other natural electrical disturbances and is propagated over the earth by ionospheric transmission. Atmospheric static is generally higher at night than in the daytime and is higher in the warm tropical areas where storms are frequent than in the colder northern regions Atmospheric static is ordinarily predominant at frequencies below a few mega

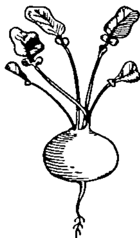
cycles, while set noise is the primary limitation at frequencies above 200 to 500 Mc In the intermediate region, the controlling noise depends on location and time of day, and may come from man made sources such as the operation of electric switches or automobile ignition noise

The very low frequency components of lightning discharges are propagated along lines of magnetic force to the antipodes, giving rise to a phenomenon known as whistlers Cosmic and solar noise are of considerable interest in the field of radio astronomy, but ordinarily are not the controlling factors in communication *See* RADIO ASTRONOMY, *see also* METEOROLOGICAL OPTICS, MICROWAVE, MICROWAVE TRANSMISSION LINES, SCATTERING (ELECTROMAGNETIC RADIATION), TRANSMISSION THEORY AND METHODS, WAVE GUIDE [K B]

Bibliography K Bullington, Radio propagation at frequencies above 30 Mc, *Proc IRE*, 35(10) 1122–1136, 1947, K Bullington, Radio propagation fundamentals *Bell System Tech J*, 36(3) 593–625, 1957, Collected papers on scatter propagation *Proc IRE*, 43(10) 1173–1526, 1955, L R O Storey, Whistlers, *Sci American*, 194(1) 34–37, 1956

Radish

A cool season annual or biennial crucifer (*Raphanus sativus*) of Chinese origin belonging to the plant order Papaverales The radish is grown for

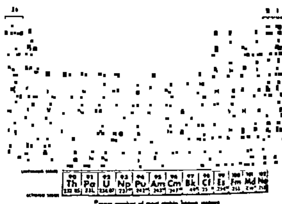


Radish, *Raphanus sativus* (From L H Bailey, ed *The Standard Cyclopedia of Horticulture*, vol 3, Macmillan 1937)

its thickened hypocotyl Propagation is by seed Varieties are classified according to root shape and season or time of maturity Colors include red yellow, white black, pink, and red white combinations Popular varieties are short season (21–25 days), Early Scarlet Globe and Comet, medium season (30–50 days), Crimson Giant, and long season or winter varieties (50–70 days), Black Spanish Commercial production, largely of the round red short season varieties, is primarily in the field but radishes are also produced commercially in greenhouses Harvesting by hand or machine begins when the roots are approximately ½–1 in in diameter, often only 21–23 days after planting *See* PAPAVERALES, VEGETABLE GROWING [H J C]

Radium

Element number 88, radium, Ra, is a rare, radioactive, chemical element which occurs in uranium minerals to the extent of 1 part for about every



3,000,000 parts of uranium Chemically, radium is an alkaline earth metal having properties quite similar to those of barium Radium is important because of its radioactive properties and is used primarily in medicine for the treatment of cancer in the manufacture of self luminous paints, in atomic energy technology for the preparation of standard sources of radiation, as a source for actinium and protactinium by neutron bombardment, and in certain metallurgical and mining industries for preparing gamma ray radiographs

Properties and uses. Biologically, radium behaves as a typical alkaline earth element concentrates in bones by replacing calcium and as a result of prolonged irradiation, causes anemia and cancerous growths The tolerance dose for the average human being has been estimated at a total of 1 microgram of radium fixed within the body

Because the radiations from radium and its decay products preferentially destroy malignant tissue radium has been used to check the growth of cancer For this use pure radium compounds are sealed in tubes or needles, or radon the gaseous decay product of radium is pumped into small tubes Radon is safer to administer and its very short half life decreases the danger of overexposure

In industrial radiography radium sulfate is generally sealed within concentric spheres or cylinders the inner one of silver and the outer of aluminum or steel In each container the quantity of radium

the thickness of metals such as copper, aluminum, brass and tin in rolling mills

The use of radium in luminescent paints for watch clock, and meter dials, and signs visible in

: depends on its α radiation striking a scin such as zinc sulfide A standard neutron for use in research, in analyses of materials tron activation and in the radio-logging of s is prepared from mixtures of radium and m Some industrial static eliminators contain

urrence Thirteen isotopes of radium are (Table 1), all are radioactive four occur lly, the rest are produced synthetically Ra^{226} is technologically important It is dis ed widely in nature usually in exceedingly quantities The most concentrated source is blende a uranium mineral occurring in large sits in the Belgian Congo, Canada and else- e and containing about 0 4 gram of radium ton of uranium Lower concentrations are d in numerous other minerals such as carnotite the Colorado plateau Detectable traces of um are found in river ocean and other natural ers

Metallurgy In the processing of pitchblende for nium radium is usually recovered with barium sulfates in an acid insoluble residue A simpli d, but typical method for the extraction of ra um from this residue is shown in the illustration fter treatment with a hot concentrated alkali car nate solution to remove sulfate and the ampho- ric metals the resulting carbonate filter cake is issolved in dilute hydrochloric acid The solution is treated first with hydrogen sulfide to remove in oluble sulfides next with ammonia to eliminate metal hydroxides and finally with sulfuric acid to reprecipitate the sulfates free of acid soluble im purities Additional purification is obtained by rep etition of the above cycle Finally the barium and

Table 1 Radium isotopes

Mass	Occurrence	Type of decay	Half life	Energy of radiation Mev
213	Synthetic	α	2 7m	6 90
219	Synthetic	α	~0 001s	8 0
220	Synthetic	α	0 03s	7 43
?21	Synthetic	α	30s	6 71
2?2	Synthetic	α	38s	6 51
223	Natural AcX	α	11 2d	5 730 (9%) 5 704 (53%) 5 596 (24%) 5 528 (9%) 5 487 (2%) 5 419 (3%)
		γ		0 026–0 44
224	Natural, ThX	α	3 64d	5 681 (95%) 5 448 (4 6%) 5 194 (0 4%)
		γ		0 241
225	Synthetic	β	14 8d	0 31
2?6	Natural Ra	α	1622y	4 ?91
		γ		0 186
227	Synthetic	β	41 2m	1 31
		γ		0 291 0 498
228	Natural, MsTh	β	6 7y	0 012
		γ		0 03
229	Synthetic	β (?)	~1m	?
230	Synthetic	β	1h	1 2

Table 2. Physical properties of radium

Atomic number	88
Atomic weight	226 05
Valence states	0 +2
Specific gravity	6 0
Melting point	~00°C*
Boiling point	~1140°C
Ionic radius, Ra^{++}	2 45 A (estimated)
Atomic parachor	~140
Decomposition potential†	1 718 volt
Heat of formation of oxide	130 kcal/mole
Magnetic susceptibility	Feebly paramagnetic

* Value of 960°C has also been reported

† Of normal solutions of radium salts with respect to the normal calomel electrode

radium are separated from each other by fractional crystallization until radium bromide of at least 90% purity is obtained

The winning of the free metal from its com ... nde has been achieved by electrolysis of a ra ...

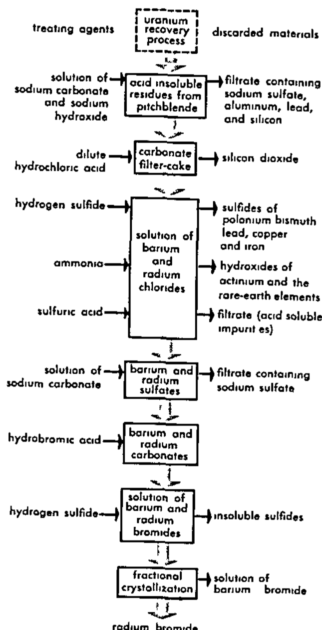
leaving a ...

When freshly prepared, radium metal has a brilliant white metallic luster Some of its physical properties are shown in Table 2 Chemically, the metal is highly reactive It blackens rapidly on ex posure to air because of the formation of a nitride Radium reacts readily with water, evolving hydro- gen and forming a soluble hydroxide

Principal compounds When first prepared, nearly all radium compounds are white, but they discolor on standing because of the intense radia tion The radiation causes a purple or brown color ation in glass on long contact with radium com pounds. Eventually the glass crystallizes and be- comes crazed Radium salts ionize the surrounding atmosphere thereby appearing to emit a blue glow, the spectrum of which consists of the band spec trum of nitrogen Radium compounds will dis- charge an electroscope fog a light shielded pho- tographic plate, and produce phosphorescence and f ...

... earths, radium halide imparts a carmine-red color to a flame

All known salts of radium are isomorphous with and similar to the corresponding barium salts Ra dium forms a sparingly soluble sulfate chromate, carbonate and iodate whereas the chloride bro- mide nitrate and hydroxide are soluble in water Most radium compounds are more insoluble in wa ter than the corresponding barium compound, the notable exceptions being the carbonate and hydrox ide A nearly specific reaction for radium is the precipitation of the nitrate from 80% nitric acid This separates radium quantitatively from all met als except barium strontium and lead Because of its chemical similarity to radium barium is fre- quently used as a carrier for minute quantities of radium The separation of these two elements is ex



Extraction of radium from pitchblende residues

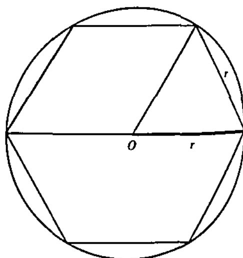
tremely difficult, and it requires fractional crystallization or precipitation, ion exchange, or solvent extraction with chelating agents The greatest separation per fractionation step is obtained with the bromides or chromates, somewhat less with chlorides and carbonates, and only slight separation with the sulfates and nitrates Complete separation of radium from barium and other metallic impurities has been accomplished by ion exchange on synthetic organic resins, but the method is not easily adapted to the separation of macro quantities of radium because the intense radiation causes the degradation of the resin which usually results in the formation of insoluble radium compounds within the ion exchange column

Analysis. Large samples of pure radium are assayed by weighing the chloride or bromide after fusion to remove water, by γ ray counting and comparison with standard samples, provided that all samples are nearly at radioactive equilibrium, or by calorimetric measurements Small samples of carrier free radium may be analyzed by α ray counting after separating or making correction for the α emitting activities of daughter elements or by use of an α ray spectrometer, an instrument which separates α rays of different energies and permits the determination of only those produced by radium Radium samples in any state of purity may be analyzed by the classical emanation method which depends on the separation and determination of radon Although the method is indirect, it is extremely sensitive *See* ALKALINE EARTH METALS NUCLEAR REACTION, RADIOACTIVE MINERALS, RADIOACTIVITY, RADON [M.L.S.]

Bibliography K W Bagnall, *Chemistry of the Rare Radioelements*, 1957, S B Schwind and F E Croxton, *Radium, A Bibliography of Unclassified Literature*, U S Atomic Energy Comm TID-363, 1950

Radius

A line segment that joins the center of a circle with any one of its points If r denotes the length of a radius, the length (circumference) of a circle



Circle with radius r, and inscribed regular hexagon

is $2\pi r$, and the area enclosed is πr^2 The radius of a circle may be marked off on the circle 6 times as a chord and a regular hexagon obtained *See* CIRCLE [L.M BL.]

Radius of gyration

A relation of the area or mass of a figure to its moment of inertia If I is the moment of inertia about a line of a figure whose area is A, the figure's radius of gyration with respect to that line is $k = +\sqrt{I/A}$ Accordingly, $I = k^2A$ For a figure of mass M, $k = +\sqrt{I/M}$, $I = k^2M$ In these equations k is measured in length units such as feet. Geometrically similar figures have equal radii of gyration about corresponding centroidal axes

If the radius of gyration of a figure with respect to an axis is k and with respect to a parallel centroidal axis is $\bar{k}$, $k^2 = \bar{k}^2 + D^2$ where D is the distance between the [illegible] INERTIA

Radome

A strong, but electrically transparent, thin shell used to house a radar antenna The shell must be large enough not to interfere with the scanning motion of the antenna In airborne radar the radome prevents the antenna from upsetting the airplane or missile's aerodynamic characteristics and protects the antenna against aerodynamic forces Shipboard radars frequently require radomes to protect ... and water damage and blast

None of the materials available for radomes possesses a dielectric constant equal to that of the atmosphere The resulting impedance mismatch causes reflections at the inner and outer faces of the shell In particular, if the shell thickness is a substantial fraction of the carrier wavelength the reflections from the inner and outer faces may

change the load on the magnetron sufficiently to pull its frequency out of the pass band of the receiver The standing wave may also distort the antenna pattern producing undesirable side lobes and changing the orientation of the main beam Furthermore the reflected power is not transmitted therefore the effective power output is reduced See TRANSMISSION THEORY AND METHODS

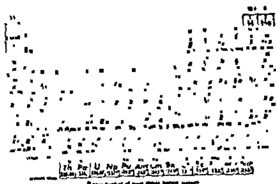
Several means are employed to prevent reflections If the shell is very thin compared to the carrier wavelength the reflections from the inner and outer faces are almost a half cycle out of phase and cancel each other This condition can be obtained easily at frequencies below and including L-band or in the uncommon situations in which a very thin weak radome can be employed Alternatively, if the shell thickness is an integral number of half wavelengths the reflections from the inner and outer faces cancel each other this arrangement is frequently employed Numerous multilayer and sandwich type radomes have been developed in the reasonably successful attempt to combine the properties needed for cancellation of reflections with mechanical strength stiffness and lightness

An additional requirement is that the radome material not cause so much loss as to subtract substantial power from the waves passing through it A number of available materials satisfy this requirement and also possess chemical and mechanical properties for ease of forming and production Most of these are organic high polymers, such as resins rubbers and fibrous material Fine glass yarn is also employed Since the usual production run is small and specialized radomes are produced either by drawing large flat sheets to the desired shape, in the case of single-layer radomes, or by low pressure molding and bonding in the case of multilayer and sandwich types See RADAR

[R I B]

Radon

The gaseous emanation produced by the radioactive decay of radium Radon is highly radioactive and decays by the emission of energetic α particles



86 and all its isotopes (symbol Rn), although the name emanation (symbol Em) is sometimes preferred for the element The element radon is the heaviest element of the noble, or inert, gas group and as such is characterized by chemical inertness All isotopes are radioactive with short half lives

Occurrence and synthesis. Radon is found in

thorium Radon (mass number 222, half life 3.82 days) occurs in the uranium radium series; thoron (mass number 220, half life 54.5 sec) is a member of the thorium family, and actinon (mass number 219 half life, 3.92 sec) is found in the actinium series All three decay by the emission of energetic α particles Thoron was discovered in 1899 by R B Owens and E. Rutherford who noted that some of the radioactivity of thorium preparations could be removed in a stream of gas. F E. Dorn discovered radon in radium preparations in 1900, and F O Giesel observed the formation of actinon in actinium compounds about 1902 The remarkable continuous formation of radioactive gaseous emanations in such samples did much to stimulate thinking on the true nature of radioactivity in the

detected in surface waters and in streams The atmosphere near the ground contains radon which has seeped from soil and rocks, all of which contain minute traces of uranium

In addition to the three natural isotopes, 14 isotopes have been synthesized by nuclear reactions of artificial transmutation in cyclotrons and linear accelerators, but none of these is as long lived as Rn^{222} The most stable of them is Rn^{211}, one half of which disintegrates in 16 hours It is certain that no isotope of radon will ever be found which is stable or long lived See NUCLEAR REACTION

Properties. Any surface exposed to Rn^{222} becomes coated with an active deposit which consists of a group of short lived daughter products, including radium A, B, C, C′, and C″ The radiations of this active deposit include energetic α-, β, and γ rays Particularly noteworthy are the penetrating γ radiations of RaC which are of practical use in radiotherapy, radiography, and other purposes In medical applications, it is common practice to keep a source of radium in some vessel from which the radon can be removed, purified, and compressed into a small glass or metal tube which is then sealed and removed Such radon tubes (also referred to as needles or seeds) may be placed near diseased tissue The penetrating radiations of the active deposit formed within the tube then irradiate and destroy the diseased tissue which typically is more radiosensitive than the surrounding normal tissue Many treatments of this type have been replaced by more successful medical treatments, but in certain cases, such as cancer of the cervix uteri this method has worked well

The ultimate decay products of radon following the rapid decay of the active deposit include radium D (lead 210), polonium and, finally, stable lead 206 Thoron and actinon also lay down an active deposit on surfaces exposed to them, but thoron, actinon, and their decay products have not had the same general interest and usefulness because of their shorter lifetimes.

When sealed into a tube filled with beryllium powder, radon has some usefulness as a laboratory source of neutrons

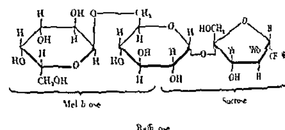
Radon possesses a particularly stable electronic configuration ($5s^2 5p^6 5d^{10} 6s^2 6p^6$, $^1S_0$) which gives it the chemical inertness characteristic of noble-gas elements Therefore, the only chemical form known is the free element Its properties are those to be expected by extrapolation from the other noble-gas elements helium, neon, argon, krypton and xenon Some physical properties are the following boiling point $-65°C$, melting point variously reported as $-71°C$ and $-113°C$, ionization potential, 10 6 electron volts (ev), ionization potential of Rn^+ ion 19 9 ev The spectrum of radon has been extensively studied and resembles that of the other inert gases Radon is readily adsorbed on charcoal, silica gel and other adsorbents, and this property can be used to separate the element from gaseous impurities The radon is desorbed from

rect gas counting of α particles in an ionization chamber, extremely minute amounts can be measured in this way Radon can also be determined by measuring in suitable counters the γ-radiation of its short lived daughters. Millicurie amounts can be estimated by comparing the strength of the γ radiation with that from a calibrated radium standard See INERT GASES, NUCLEAR RADIATION (BIOLOGY), RADIOACTIVITY [E K HY]

Bibliography K W Bagnall, *Chemistry of the Rare Radioelements*, 1957

Raffinose

The best known trisaccharide (oligosaccharide) widely distributed in higher plants. Sugar beets contain about 0 5% of this sugar The best known sources are cottonseed meal and the manna of *Eucalyptus* It is also known as melitose, melitriose, gossypose, and *O* α D galactopyranosyl (1→6) *O* α D glucopyranosyl (1→2) β D-fructofuranoside This trisaccharide is nonreducing, it crystallizes from aqueous alcohol or acetic acid as the pentahydrate, melting point 80°C, $[\alpha]_D^{20}$ +105 2° (in water) *See* OLIGOSACCHARIDE, OPTICAL ACTIVITY.



Raffinose

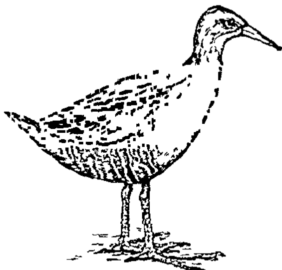
Complete acid hydrolysis gives 1 mole each of D galactose, D glucose, and D fructose In structure it comprises melibiose and sucrose with the central D glucose in common (*see* FRUCTOSE, GALACTOSE, GLUCOSE)

Raffinose can be hydrolyzed by enzymes in two ways Invertase (β D fructofuranoside) hydrolyzes

melibiose residue to yield D galactose and sucrose. [W Z H]

Rail

Any of several species of the family Rallidae, a group which also includes the coots and gallinules. There are seven rails in the United States They are small to moderate sized wading birds living in marshes and in weedy pond and lake borders Rails



The King rail, Rallus elegans; length to 19 in (from E L Palmer Fieldbook of Natural History, McGraw-Hill, 1949)

brown or gray streaked birds with remarkable ...ealing coloration, they rely heavily upon this ...ection to escape enemies because they are ...ble only of a feeble, fluttering flight for a ... hard pressed They have large,

... the rails are listed as game birds ... limited amount of shooting. See COOT, GALLINULE, GRUIFORMES [J D B]

...ilroad engineering

...branch of transportation engineering concerned ...

...tensions of line though this is a trend ... reaches back even to the turn of the century Construction of new passenger stations and terminals is primarily a thing of the past Many of the new engine terminals built during World War I and shortly thereafter have been abandoned as diesel power replaces steam power Many water stations and water pans have been given up Such developments have led to more intensive use of existing facilities and to the abandonment or restricted operation of unprofitable lines Extensions of line have been primarily confined to new properties or industrial developments particularly in countries outside the United States

The modern railroad engineer must be a keen student of economics for among the more difficult problems he faces are those involving economics as well as engineering One of these involves mass commuter transportation into large centers of population Another concerns public demand for new and elaborate passenger stations and new union stations in population centers despite diminishing passenger traffic and revenue A third deals with consolidation of railroads and abandonment of nonproductive lines

The expansion of the highway system of the United States has required engineering departments to give a great deal of attention to problems created by highway railroad crossings. The problems are not only in the design of such crossings but also in the financing which is shared with governmental bodies and consequently involves extensive negotiations at the local, state and Federal levels

In addition the modern railroad engineer is concerned with improvement of signal systems motive power rolling stock and yard facilities Recent developments and trends in these areas will be discussed in the following paragraphs

Motive power The primary motive power unit in service today is the diesel engine The steam locomotive has all but disappeared and the present ... ation of electrical operation because of high initial costs In replacing the steam locomotive the modern diesel has lowered operating and maintenance costs and reduced the number of locomotives necessary to maintain adequate service A number of diesel units can be coupled together and operated from one cab The number usually so linked depends upon the type of train and the grades to be encountered

Power is developed by a diesel engine driving a generator, which in turn feeds electric traction motors mounted on each axle This type of power transmission is easy to control and accounts for the efficiency of the diesel at all speeds under most load conditions.

Railroads have obtained greater use from diesels than from steam or electric units because the diesel can perform many different jobs efficiently and spends less time in the shop for maintenance, both routine and major Diesel maintenance is less time-consuming because there are many interchangeable parts—whole motor generator sets can be replaced and a unit quickly restored to service

The type of diesel most frequently used is the general purpose road switcher It has been successful on all railroads, hauling fast passenger trains, long haul freights, and local commuter trains, and is also used in both local and interchange yards Other types are designed especially for fast passenger or freight service and differ primarily in gear ratios and horsepower ratings A more recent development is the gas turbine locomotive which is in limited freight service in the West It is basically like the diesel except that the generator is driven by a gas turbine fired by liquefied petroleum gas. This locomotive may succeed the diesel but for the present must be considered experimental Railroads and coal companies have also considered using coal burning steam turbines but an economically practical design has not been developed

Rolling stock. Many new kinds of freight car have been developed in recent years Among them are the insulated boxcar with underframe heater to keep perishables from freezing, mechanical refrigerators to replace ice filled types large capacity hopper and tank cars for solid and liquid bulk shipments, and cars with built in movable compartments or walls to reduce damage due to load shifting Many new and existing cars have been fitted with cushion underframes to absorb the sudden shocks of starting and stopping Cars also have been equipped with improved wheel bearings to permit higher speeds, better rolling characteristics and a reduction in maintenance. One notable recent system of freight hauling is the piggy back train truck trailers are loaded on flat cars and carried between terminals Special cars with hold down equipment are being built for this purpose

Signaling. Signal development has brought far reaching improvement in railroad operations Most important is centralized train control, which enables a single tower to control all switches and signals over several hundred miles of mainline

track Train movements, signals, and switches are indicated on a master board by colored lights The operator can schedule meetings of trains at long passing sidings so that a train need not be side tracked until another passes In many cases one track with centralized train control is more efficient than two tracks without it

With two other developments, automatic train control and speed control a train running through a stop signal or exceeding a speed limit can be electronically stopped or slowed, eliminating the possibility of human failure

Another development is reverse signaling, which is generally used only on multiple track sections with heavy train traffic in one direction at one time of day and in the other at another time An example is a passenger terminal handling commuter traffic Extra signals are installed so that trains traveling in either direction can be regulated Therefore tracks can be used according to demand, three tracks of a 4-track mainline can be set for travel in one direction in the morning and for travel in the opposite direction in the evening

The system most commonly used today is the automatic block in which a signal guards the section of track immediately beyond it and indicates whether another train is present This system works in one direction only, and if simultaneous 2-way operation is necessary, two tracks must be used Of course, automatic safety control devices can be incorporated into the block system

An important device which can be used with any signaling system is a miniature signal board in the locomotive cab It shows how the next signal is set and serves as an advance warning and double check for the engineer, especially in bad weather

Communications. Two way radio has done much to expedite freight train operations Many locomotives, cabooses, and towers are so equipped enabling train conductors to talk to the engineer or to tower operators to receive orders while en route Portable and fixed radio installations are used in large yards for contact between car inspectors tower operators, locomotive engineers and other yard personnel

Yard operation. One of the more expensive elements of a railroad system is the classification yard, where freight cars from incoming trains are sorted and combined by destination into new trains (Fig 1) These yards are either at a level

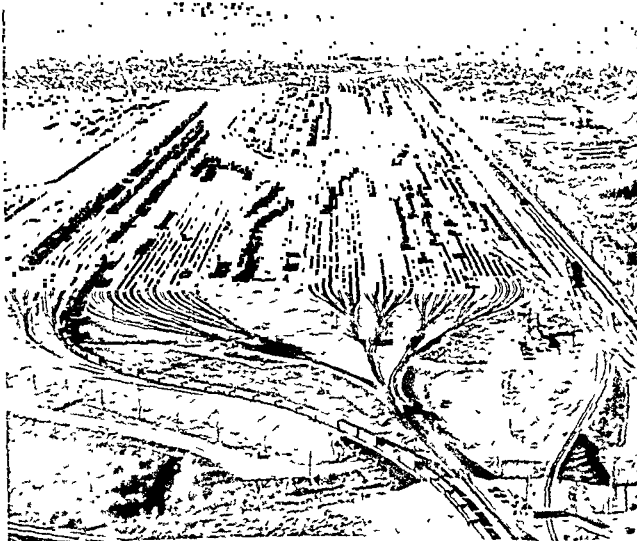


Fig 1 The Englewood classification yard at Houston Texas With the assistance of many electronic devices, including a computer, the yard can pass up to 3800 cars a day over the hump (foreground) (From Southern Pacific Lines)

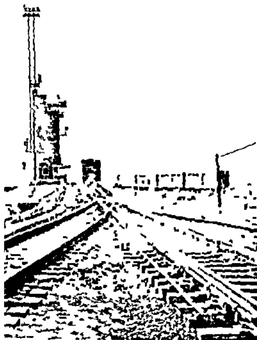


Fig 2 The hump as seen from classification tracks within the Cicero Illinois yard (From Burlington Lines)

grade and cars must be pushed to the proper track ... h engine or they are on a slope and cars

be made up into new trains

At one time it was necessary for men to ride each car down the grade and brake it to a safe speed for coupling This time and manpower consuming function has now been assumed in the larger modern yards by automatic devices called car retarders (Fig 3) Several of these are usually located at points between the hump and the classification tracks The retarder installed along the inner faces of both rails operates by pressing against the wheels of a rolling car slowing it to a speed safe for coupling An operator controls the retarder by turning a knob in the yard tower This device has expedited operations reduced time loss and improved safety Other devices have also been developed to aid control of yard operations from the tower Automatic scales record the weight of the car as it passes the hump closed circuit television is used to report the railroad name and number as well as to permit visual inspection of the underframe and radar units set between rails are used to indicate the rollability speed and distance the car must travel to coupling With careful planning of grades and location railroads have combined ... ds into one

Construction standards Each railroad has its standard requirements for alignment grade track construction procedures and other details. Many of these standards are based upon the *Manual* and *Proceedings of the American Railway Engineering Association* (AREA)

Alignment and grades On main lines where maximum speed and tonnage are objectives curvatures are usually no greater than approximately 1°30′ (2°00′ at most) and grades not in excess of 0.5% The degree of curve is defined as the angle at the center of a simple circular curve subtended by a 100-ft chord The rate of grade is the vertical distance in feet that a grade line rises or falls in 100 ft It is usually expressed as a percentage

Branch lines and industrial and sidetrack extensions ordinarily have curves no greater than 6–8° and grades not over 3% If a limited number of cars is to be handled on a sidetrack sharper curves and steeper grades are sometimes employed

Road steam locomotives are designed to handle curves of up to 18°30′ or 19° Diesels will negotiate curves of 20–40°

Roadbed Although there is variation among the railroads side slopes generally are built to have a 1.5 1 ratio between the horizontal and vertical dimensions Local conditions and the character of the ground are also determining factors

Track spacing Track spacing has varied in the past from 12 to 13 ft The present tendency is to provide 14 ft in main track construction and greater space for subsidiary tracks

Clearances The requirement for overhead clearance is usually 22 ft Operation has been possible in the past with 16–18 ft. Clearance of 18–19 ft is required as a minimum in connection with electric overhead operation but if clearance is less than 22 ft bridge warning signs (ticklers) must be installed These requirements are constantly be

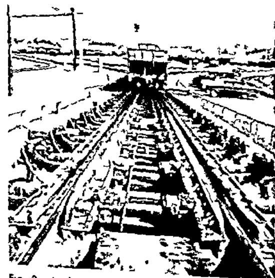


Fig 3 Looking down the throat of the master retarder at the Boyles yard Birmingham Alabama (From Louisville and Nashville Railroad)

coming more rigid and are fixed by law in a number of states

Ballast A fundamental of good railroad track construction is the provision of 12 in subballast made of gravel, rock, cinders, furnace slag, or any material that will supply a fairly porous but substantial base On it is placed a 12 in layer of good limestone or other rock crushed to about 1¾ in size This layer is carefully surfaced, the ties and rails are laid, bolted, and spiked, and the ties are tamped to bring the top of rail to exact surface and alignment Power operated track tools have done much to improve maintenance and to reduce cost (Fig 4)



Fig 4 With one operator this automatic spike driver does the work that formerly required nine men It straightens each tie and has four pneumatic hammers that drive all four spikes either simultaneously or in any combination (*From W H Nichols & Co*)

Crossties Crossties of a large variety of woods are used by railroads in the United States All are pressure creosoted Except for special uses, cross ties for high grade track are usually 7 in by 9 in by 8 ft 6 in Bridge ties are usually 8 in by 8 in by 9 ft Switch timbers of good quality are used Their dimensions are 7 in by 9 in by the various lengths required for turnouts

Tie plates Rolled steel tie plates are generally used They are ordinarily 10–16 in in length about 7½ in wide with shoulders, and carry holes for spikes

Rail Each railroad has its own standard for rail sections There are many different weights but sections for mainline traffic generally weigh 127–140 lb/yard Some railroads handling particularly heavy coal and ore tonnages use 153-lb rail Many roads have adopted the AREA standard of 133-lb rail for mainline operation and 112–115 lb rail for sideline and sidetrack A significant development is the use of welded rail sections sometimes called ribbonrail Standard 39-ft lengths of rail are welded into continuous strips before installation The strips, usually under ¼ mile in length, are transported in open end gondolas As welded rail does not have the usual joints a major source of wheel damage is eliminated Generally welded sections have been used where there is extremely heavy traffic or where maintenance is especially difficult, as in tunnels, through-station platforms highway railroad crossings, and tracks in streets

Switches and frogs The present standard switch point is called the split switch Switch rails are 11–45 ft long Curved switches and frogs permit a smoother movement on high speed track. A turnout frog is officially defined as "A track structure used at the intersection of two running rails to provide support for wheels, and passageway for their flanges, thus permitting wheels on either rail to cross the other" Frogs may be either bolted or rigid or spring rail frogs The rigid frog is used principally in yard turnouts, the spring frog is used in nearly all main track turnouts and in many other tracks where one of the turnout rails carries the predominant traffic

Crossings Crossings of one track with another with crossing angles of approximately 10-90° are subjected to both heavy and light traffic traveling at high and low speeds Like turnout frogs crossing frogs are built with a variety of construction techniques and materials They are placed in [illegible]

Where two railroads cross one with much more traffic than the other, the former often will be favored with a stronger section As replacement or even heavy repair of a crossing nearly always affects train operation, every effort is made to provide equipment of great durability

Track gage A 4-ft 8½-in gage has been universally adopted in the United States and to a great extent abroad Many of the countries with a variety of gages are changing to a more uniform practice A few railroads in the United States are using a 3 ft gage in lumber regions and mountainous country

Superelevation The vertical distance of the outer rail above the inner rail is called its superelevation Tracks are superelevated on curves so that cars and locomotive will be balanced against the centrifugal force and will operate smoothly The maximum superelevation usually is about 6 in Safety and comfort limit the speed at which a passenger train may negotiate a curve any speed that gives comfortable riding on a curve may be considered safe In planning superelevation of a track the speed at which traffic will be moving is the first element to be determined On curves where the speed will vary superelevation must be based on a compromise that suits the track If only one class of service is to use the track superelevation should not exceed about 7 in If the traffic will be mixed it should not exceed about 6 in

Sidings Passing sidings used regularly by freight and passenger trains require more durability than the ordinary light service industrial track If they are frequently used by the railroad passenger [illegible] require a high standa[illegible]

rial tracks are usually owned and maintained industry and do not require as much main
[R E D]
:e. Amer can Railway Engineering

Dick (ed) *Railway Track and Struc u es* pedia 8th ed 1955 W W Hay *Railroad* eering 1953 A M Wellington *The Eco* *e Theory of the Location of Railways* 1887

n shadow

rea of diminished precipitation on the lee side ountains There are marked *rain shadows* for nple, east of the coastal ranges of Washington gon and California, and over a larger region h of it and east of the Cascade Range and ra Nevadas Precipitation on the northern Ore- coast is around 100 in per year east of the stal ranges in the Willamette Valley it is about in and east of the Cascades it drops to around in All mountains decrease precipitation on their but rain shadows as is shown by annual to- s are sometimes not marked if moist air comes quently from different directions as in the Ap- lachian region

The causes of rain shadow are (1) precipitation much of the moisture when air is forced upward the windward side of the mountains (2) deflec (3) down rms the air

iu uwe s w TMOSPHERIC DIABATIC CHANGE CHINOOK PRECIPITATION (ME- EOROLOGY) [J R F]

Rainbow

Colored arcs seen in the skies when the sun or moon is illuminating large numbers of falling rain- drops Such arcs are centered around the antisolar point (180° from the sun) Among the parallel rays striking a water droplet there is one ray which after one internal reflection leaves the drop at the smallest angle of deviation from the direc tion of the incident rays All other rays emerge from the drop after an internal reflection at larger angles of deviation The ray of minimum deviation also called Descartes ray has the greatest inten sity and thus is more visible than the others Since the angle of deviation for such a ray of minimum dev ation is smaller for the red l ght and larger for the blue the rays of minimum deviation from a large n mber of water drops of the same size are seen as arcs of different colors red on the upper part, and blue on the lower part of the bow The angular distance of these arcs from the antisolar po nt is greater than 42° and the system of arcs is called the primary rainbow The rays of minimum deviations, after two internal reflections form simi lar colored arcs at the angular distance of 51° from the antisolar point to make the secondary rainbow In this rainbow the color sequence is reversed with the red arcs visible on the lower part the blue on pper part of the bow [z s]

Rajiformes

The skates and rays constitute one of the two Re- cent orders of the subclass Elasmobranchii The order is also known as the Batoidea The skates and rays differ from the sharks (order Squali formes) in having ventral gill slits the edge of the pectoral fin attached to the side of the head ante- rior to the gill clefts, and the upper margin of the orbit not free from the eyeball Guitarfishes the most generalized rays appeared in the Upper Ju rassic, but most major groups of rays date from the Cretaceous



Atlant c sting ray Dasyatis sab na (After G B Goode Great International F sher es Exh bit on London 1883 U S Nat'l Museum Bull 27 1884)

Rays have been classified in 5 suborders 16 fam ilies 47 genera and 300–350 species They occur in all oceans and most are sluggish bottom inhab- itants although the mantas commonly swim at the surface A few especially sting rays and sawfishes penetrate estuaries of tropical rivers or even live far upstream Rays vary in size from less than 1 ft

The torpedoes have a pair of enlarged electric or gans located lateral to the eyes on the expansive pectoral fins these can deliver a strong and tem porarily disabling shock to bathers More to be feared however are the sting rays which lie in shal low water and if stepped on are able to inflict a dangerous or fatal wound with the serrated tail spine and its venom gland

Internal fertilization is practiced by all rays with the modified pelvic fins serving as intromittent organs Most rays are ovoviviparous retaining the eggs in the oviducts until birth but skates deposit eggs that are protected by horny cases Rays are carnivorous feeding on a wide variety of marine worms mollusks crustaceans and other inverte- brates as well as small fishes.

Although rays are abundant in many seas they are of minor commercial importance *See* BATOIDEA FOSSILS ELASMOBRANCHII [R M B]

Bibliography J Tee Van et al (eds.) *Fishes of the Western North Atlantic* Sears Foundation for Marine Research Mem 1 pt 2 1954

Raman effect

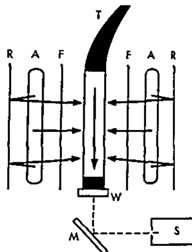


Fig 1 Schematic diagram of arrangement for exciting and observing the Raman effect

A phenomenon observed in the scattering of light on passage through a material medium, whereby the light suffers a change in frequency and a random alteration in phase Raman scattering differs in both these respects from Rayleigh and Tyndall scattering, in which the scattered light has the same frequency as the unscattered and bears a definite phase relation to it The intensity of Raman scattering is roughly one thousandth that of Rayleigh scattering in liquids and smaller still in gases For an extended discussion of Rayleigh scattering, *see* SCATTERING (ELECTROMAGNETIC RADIATION) *See also* TYNDALL EFFECT

Discovery. Because of its low intensity, the Raman effect was not discovered until 1928, although the scattering of light by transparent solids, liquids, and gases had been investigated for many years prior to that time Prompted by A H Compton's observation of frequency changes in x rays scattered by electrons (Compton effect), the Indian physicists C V Raman and K S Krishnan examined sunlight scattered by a number of liquids With the help of complementary filters they found that there were frequencies present in the scattered light which were lower than the frequencies in the filtered sunlight They then showed, by using light of a single frequency from a mercury arc, that the new frequencies present in the scattered radiation were characteristic of the scattering medium Within a few months of Raman and Krishnan's first announcement of their discovery the Russian physicists G Landsberg and L Mandelstam communicated their independent discovery of the existence of the effect in crystals In Russian literature *the phenomenon is now referred to as combi*

ring is ana
ction of new

frequencies appearing in the spectrum of monochromatic radiation scattered by a substance is characteristic of the substance and is called its Raman spectrum Although the Raman effect can occur in atomic spectra it is of greatest interest in molecular spectra The present technique of Raman spectroscopy which was originally developed by the American physicist R W Wood about 6 months after Raman and Krishnan's discovery, is shown in schematic form in Fig 1

A number of arcs A usually containing mercury vapor as the light producing substance are surrounded by appropriate reflectors R to increase the intensity of available radiation Light from the arcs passes through the filters F which transmit one of the several monochromatic frequencies of the mercury radiation and which in principle absorb all other frequencies The monochromatic radiation then enters the cylindrical tube T containing the liquid or vapor to be studied This tube is horn shaped and blackened at one end so that when viewed through the window W at the other end the tube presents a perfectly black background against which the scattering is seen A representative tube for the study of liquids would hold from 5 to 50 ml and would be 10–15 cm long and 1–2 cm in diameter Solid materials can also be studied if optically clear samples of large enough size are obtainable

Raman scattering is more or less uniform in all directions and the scattering is studied at right angles to the incoming radiation In this way the strong exciting light interferes least with the observation of the scattered light Light scattered in the direction of the vertical arrow (Fig 1) passes out the window W and is sent to the spectrograph at S by a suitable optical system represented by the mirror M The spectrograph resolves the Raman frequencies from the Rayleigh scattered exciting frequency (the so-called Rayleigh line) and focuses them on a photographic plate

Prior to the invention of the photomultiplier

possible to record Raman spectra automatically thereby increasing the speed and precision of measurement of intensities of Raman frequencies The appearance of a photoelectrically recorded Raman spectrum is sketched in Fig 2 Intensity of the scattered light is plotted vertically and frequency in wave number units horizontally

Raman spectroscopy is of considerable value in determining molecular structure and in chemical analysis Molecular rotational and vibrational frequencies can be determined directly, and from these it is sometimes possible to evaluate the molecular geometry or at least to find the molecular symmetry The procedures for doing this are indicated briefly in another article (*see* MOLECULAR STRUCTURE AND SPECTRA) and are described in greater detail in the books cited in the bibliography Even when a precise determination of structure is not possible much can often be said about the arrangement of atoms in a molecule from empirical information about the characteristic Raman frequencies of groups of atoms This kind of information is closely similar to that provided by infrared

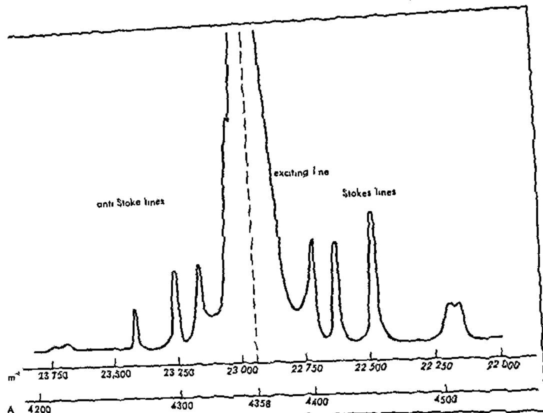


Fig 2 Sketch of the Raman spectrum of carbon tetrachloride as recorded photoelectrically Intensity of the light in arbitrary units is recorded vertically against horizontal wave number scale (cm⁻¹) (frequency in cycles/sec = wave numbers times velocity of light in cm/sec) The lower scale shows wavelengths in angstroms The exciting line (Rayleigh line), which is scattered unchanged in frequency, is the mercury vapor blue line at 4358 A (22,938 cm⁻¹ vac) The so-called Stokes lines appear at lower frequencies and the less intense anti-Stokes lines at higher frequencies than that of the exciting radiation

spectroscopy and in fact Raman and infrared spectra often provide complementary data about molecular structure *See* INFRARED SPECTROSCOPY

Theory The mechanism of the Raman effect can be envisaged either by the corpuscular picture of light or from the point of view of the wave theory Both pictures merge in the basic quantum theory

through a material ...

with atoms or molecules If the collision is elastic, the photons bounce off the molecules with unchanged energy E and momentum and hence with unchanged frequency ν Such a process gives rise to Rayleigh scattering If the collision is inelastic, the photons may *gain energy* from, or lose it to the molecules. A change ΔE in the photon energy by Planck's relationship $E = h\nu$ must produce a change in the frequency $\Delta\nu = \Delta E/h$ Such inelastic collisions are rare compared to the elastic *ones and* the Raman effect is correspondingly much *weaker than Rayleigh* scattering

In the wave picture of the effect, the electromagnetic waves which constitute the incoming monochromatic radiation sweep through the material medium *Since* the atoms and molecules composing the medium are made up of negatively charged electrons and positively charged nuclei the electric field of the light waves sets the electrons to oscillating, chiefly with the frequency of the incoming

...

passing ...

This process is analogous to the elastic collisions given by the corpuscular picture

The ability of the electrons and nuclei in a molecule to be displaced by an electric field is called the molecular polarizability α It is not a simple property of the molecule, but depends in a complicated way on the frequency of the electric field on the molecule's orientation, and on the internal motions of the nuclei and electrons Thus α varies periodically with molecular rotation and vibration, and thereby the effect of a light wave on the electrons and nuclei of a molecule can be changed

When a monochromatic *light wave sweeps* through a transparent medium containing rotating and vibrating molecules most of the wave is re-created unchanged by the oscillating electrons, but because of the periodic changes produced in α by rotation *and vibration, new frequencies* are added to the light wave The appearance of these new frequen

cies, whose values are determined by the rotational and vibrational energies of the molecules, is analogous to the result of the inelastic collisions of the corpuscular model For the wave picture of the Raman effect, the quantity α is the basic quantity The intensity of the Raman effect depends on the magnitude of the changes produced in α by molecular rotation and vibration and the number and values of new frequencies (usually expressed as frequency shifts $\Delta\nu$ from the original monochromatic frequency) depend on the variation of α with the frequencies of rotation and vibration

The temperature of the scattering molecules is an additional factor which affects the intensity of Raman frequencies higher than the exciting frequency (the so called anti Stokes lines, see Fig 2) The anti Stokes lines, having higher frequencies, correspond to photons which have higher energy than those of the exciting light, and this energy must come from the molecules If the molecules do not have any available vibrational or rotational energy, that is, if they are at the absolute zero of temperature, there is no possibility of inelastic collisions in which energy is transferred from a molecule to a photon Hence anti Stokes lines vanish at absolute zero At nonzero temperatures the intensity ratio of an anti Stokes line to a Stokes line is given to a good approximation by the ratio of the number of molecules which can give up the corresponding energy to the number which can accept it from the light wave [R C L]

Bibliography G R Harrison, R C Lord and J R Loofbourow *Practical Spectroscopy*, 1948, G Herzberg *Infrared and Raman Spectra of Polyatomic Molecules* vol 2 2d ed 1945, E B Wilson J C Decius, and P C Cross *Molecular Vibrations*, 1955

Ramie

The plant *Boehmeria nivea*, of the nettle family Urticaceae, is the source of a natural woody fiber resembling flax It has an attractive silky luster of fine quality (Fig 1) Ramie also called rhea and China grass is the toughest longest, strongest and most durable fiber known It can be bleached to extreme whiteness It is more resistant to mildew than many plant fibers and when subjected to moisture its strength increases

The plant is a coarse herb or half shrub 4–7 ft tall and the fibers are in the phloem (food conducting tissue) It is grown extensively in China Japan, Formosa, India, and the Philippine Islands and to some extent in southern Europe Introduced into the southern United States ramie made excellent growth, but the costly hand labor required to extract the fibers and remove their coating of gum made production unprofitable Modern machinery designed to do this work gives promise of more economical production

When combed, ramie has half the weight of linen, but is much stronger, coarser, and more absorbent It has a permanent luster and good affinity for dyes Ramie is used as filling yarn in mixed



Fig 1 Ramie, *Boehmeria nivea* Plant height 3–8 feet (From P DeJanville, *Atlas de Poche des Plants Utiles des Pays Chauds, Librairie des Sciences Naturelles, 1902*)

woolen fabrics, as adulteration with silk fibers, and as a substitute for flax The China grass cloth used in the Orient is made of ramie This fiber is also useful for rope, twine, and nets *See* FIBER, NATURAL, FLAX, LINEN, PHLOEM, SILK, TEXTILE, URTICALES, WOOL. [M D P, P.D.S]

Ramie diseases. *Phoma boehmeria* causes diffused stem lesions and bending of the stalks of ramie Both stems and leaves are attacked by *Colletotrichum boehmeriae*, which also stains the ramie fiber *See* LEAF (BOTANY), STEM (BOTANY) In Florida *Rhizoctonia solani* has destroyed fields of young plants and caused a breaking of the stalks of older plants at the point of infection (Fig 2) Leaf spots caused by *Cercospora boehmeriae* and *C krugiana* retard the growth of young plants and may defoliate plants of poor vigor (Fig 3) *Rosellinia* sp is a root pathogen in Japan and the



Fig 2 Young ramie plants attacked by *Rhizoctonia solani*



Fig 3 Leaf spot of ram e caused by *Cercospora boehmeriae*

Philippines *See* ROOT (BOTANY) Other fungi at tacking ramie are *Ascochyta boehmeriae*, *Fomes noxius* and *Fusarium* sp

Cercospora leaf spot may be controlled by dust ing or spraying diseased plants with the fungicide Dithane (sodium ethylenebisdithiocarbamate)

[T E SU]

Ramjet

The simplest of the air breathing propulsion en gines (Fig 1) In flight air enters the front of the diffuser at high velocity The diffuser is shaped to reduce the air speed and hence its kinetic energy as it passes through With an efficient diffuser the reduction in kinetic energy results in a nearly equal increase in potential energy in the form of an increase in air pressure This higher pressure air enters the combustion chamber where fuel is continuously injected and burned The hot gas is then ejected rearwardly through the discharge nozzle at velocity V_J greater than flight speed V_0 To a first approximation thrust F is

$$F = M(V_J - V_0)$$

where M is the mass of air per second flowing through the engine

Characteristics The objective of the ramjet cy cle is to provide a jet velocity V_J that is consider ably greater than the initial velocity V_0 This in crease in air velocity represents an increase in kinetic energy The efficiency of a ramjet in con verting the chemical energy in the fuel into kinetic energy of the air stream depends upon the ratio of the pressure in the combustion chamber to the ambient air pressure This pressure ratio in turn depends upon the flight speed or, more exactly, up

increase

efficiency of the ramjet is zero Thus the ramjet has no thrust at take off As the flight speed in creases the pressure ratio and hence efficiency in crease which causes an increase in thrust and a reduction in specific fuel consumption (Fig 3)

At any given flight Mach number, maximum thrust is developed when sufficient fuel is injected into the combustion chamber to consume substan tially all of the oxygen in the air passing through This represents the largest amount of heat that can be introduced into the air Greater efficiency of utilization of the fuel however, is obtained when less than the maximum burnable amount is in jected The efficiency of utilization of the fuel is represented by the specific fuel consumption (pounds of fuel consumed per hour per pound of thrust) The higher the efficiency, the lower is the specific fuel consumption Curves in Fig 3 are for

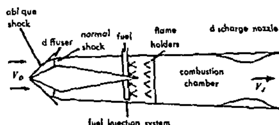


Fig 1 Diagram of a ramjet engine

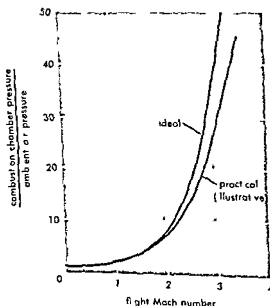


Fig 2 *Effect of flight Mach* number on pressure ratio across the inlet diffuser

cies, whose values are determined by the rotational and vibrational energies of the molecules, is analogous to the result of the inelastic collisions of the corpuscular model For the wave picture of the Raman effect, the quantity α is the basic quantity The intensity of the Raman effect depends on the magnitude of the changes produced in α by molecular rotation and vibration, and the number and values of new frequencies (usually expressed as frequency shifts $\Delta\nu$ from the original monochromatic frequency) depend on the variation of α with the frequencies of rotation and vibration

The temperature of the scattering molecules is an additional factor which affects the intensity of Raman frequencies higher than the exciting frequency (the so called anti Stokes lines, see Fig 2) *The anti Stokes lines, having higher frequencies,* correspond to photons which have higher energy than those of the exciting light, and this energy must come from the molecules If the molecules do not have any available vibrational or rotational energy, that is, if they are at the absolute zero of temperature, there is no possibility of inelastic collisions in which energy is transferred from a molecule to a photon Hence anti Stokes lines vanish at absolute zero At nonzero temperatures the intensity ratio of an anti Stokes line to a Stokes line is given to a good approximation by the ratio of the number of molecules which can give up the corresponding energy to the number which can accept it from the light wave [R C L]

Bibliography G R Harrison R C Lord, and J R Loofbourow, *Practical Spectroscopy*, 1948, G Herzberg *Infrared and Raman Spectra of Polyatomic Molecules*, vol 2 2d ed, 1945, E B Wilson J C Decius, and P C Cross, *Molecular Vibrations* 1955

Ramie

The plant, *Boehmeria nivea*, of the nettle family Urticaceae, is the source of a natural woody fiber resembling flax It has an attractive silky luster of fine quality (Fig 1) Ramie also called rhea and China grass is the toughest longest, strongest and most durable fiber known It can be bleached to extreme whiteness It is more resistant to mildew than many plant fibers and when subjected to moisture, its strength increases

The plant is a coarse herb or half shrub 4–7 ft tall and the fibers are in the phloem (food conducting tissue) It is grown extensively in China Japan Formosa, India, and the Philippine Islands and to some extent in southern Europe Introduced into the southern United States ramie made excellent growth, but the costly hand labor required to extract the fibers and remove their coating of gum made production unprofitable Modern machinery designed to do this work gives promise of more economical production

When combed ramie has half the weight of linen, but is much stronger, coarser, and more absorbent It has a permanent luster and good affinity for dyes Ramie is used as filling yarn in mixed



Fig 1 Ramie, *Boehmeria nivea* Plant height 3-8 feet (*From P DeJanville, Atlas de Poche des Plants Utiles des Pays Chauds, Librairie des Sciences Naturelles, 1902*)

woolen fabrics, as adulteration with silk fibers, and as a substitute for flax The China grass cloth used in the Orient is made of ramie This fiber is also useful for rope, twine, and nets See FIBER, NATURAL, FLAX, LINEN, PHLOEM, SILK, TEXTILE, URTICALES, WOOL [M D P, P D S]

Ramie diseases. *Phoma boehmeria* causes diffused stem lesions and bending of the stalks of ramie Both stems and leaves are attacked by *Colletotrichum boehmeriae*, which also stains the ramie fiber See LEAF (BOTANY), STEM (BOTANY) In Florida, *Rhizoctonia solani* has destroyed fields of young plants and caused a breaking of the stalks of older plants at the point of infection (Fig 2) Leaf spots caused by *Cercospora boehmeriae* and *C krugiana* retard the growth of young plants and may defoliate plants of poor vigor (Fig 3) *Rosellinia* sp is a root pathogen in Japan and the



Fig 2 Young ramie plants attacked by *Rhizoctonia solani*

Range land conservation

The major purpose of range land conservation is to secure maxim im forage production for each site consistent with ecological stability of the vegeta tion and of the soil (*see* ECOLOGY SOIL) Also where applicable range land conservation is con cerned with the preservation of watersheds with timber production and with recreation Range con servation is one of the youngest fields of natural resource management

Grazing land economics The public range is an integral part of many livestock operations in the West where approximately 10 000 000 head of live stock receive about one-third of their annual for age requirements from public lands Consequently the publ c range is an important element in the na tional production of meat wool, and leather Gov ernment receipts from grazing permits forest prod ucts, and mineral leasing on the public range to taled about $65 000 000 in 1952 over five times

of the

r yield and reduces erosion and downstream

on

Types of grazing lands Artificial pastures and ie tall grass prairies (now largely in corn and heat) are not considered range land The mid rass and short grass plains west of the prairies nd east of the Rocky Mountains the arid and emiarid grasslands of the Southwest and the Great Basin the grasslands of intermountain valleys, the brush and open woodlands of the mountains and montane and subalpine meadows are all parts of the western range of the United States Depending on the local climate some of it provides only win ter grazing whereas other parts especially at higher elevations provide from a few weeks to a few

one tenth of the United States Previous to the Taylor Grazing Act (June 28 1934) western graz ing lands of about 170 000 000 acres were open to free access and use without public control and management Now 156 000 000 acres are organized into grazing districts 3 000 000–9 000 000 acres in extent that are administered to conserve and regu late the p ibl c grazing land and to help stabilize the li estock industry The p iblic range is used by about 20 000 private stockmen under a system of perm ts and a code that seeks to guarantee proper use of the range and return to the government fair compensation for use

A long history of cutthroat competitive grazing and of conflict between cattle and sheep operators and between them and agricultural settlers had re sulted in widespread deterioration of the range Productivity which was naturally low in the more regions (averaging about 70 lb of air dry for age per acre on the public range in contrast to about 2 800 lb per acre for average hay land in the United States) was reduced by overgrazing to ng capacity in many places

were reduced in amount and weeus a desirable species were increased This is well illus trated in Texas, New Mexico Arizona and south western California by mesquite which now occupies about 50 000,000 acres (twice what it did 50 years ago) and which cuts grass forage to one-third or less of full production when the mesquite bushes in per acre In the West over

the passage of the Taylor Grazing Act has many ac complishments to its credit increase of range for age more livestock products greater protection of public lands more usable water less erosion and downstream sedimentation more flood control more wildlife and more recreational opportunities yet the Bureau estimates that in a quarter of a century not more than 10% of the needed range improve ment work has been done largely because of inade quate Federal appropriations The Bureau s pro gram includes range inventories and management plans improved fire protection watershed treat ment works (regrassing water conservation and erosion control dams water spreading structures) pest and rodent control and range management im provements (stock water developments range fenc ing corrals livestock and truck trails) In spite of this program about 50% of Federal range lands are still in a state of severe to critical erosion 32% are suffering moderate erosion and only 18% are in a condition of unaccelerated or no erosion

Range inventory The problem is to determine what the vegetation soil and climate will permit with regard to forage production as measured by the land s animal carrying capacity The range in ventory includes a quantitative evaluation of the vegetation its palatability the nutritional value of each species the status of the vegetation in the plant successional process, and the time and de gree of permissible use



Cattle on a western summer range

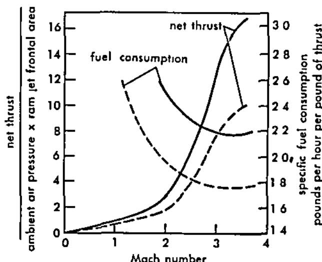


Fig 3 Effect of flight Mach number on thrust and specific fuel consumption of ramjet engine Solid lines are for fuel-air ratio adjusted for maximum thrust, broken lines are for fuel air ratio adjusted for maximum efficiency

the two operating conditions of maximum power and maximum efficiency

Combustion. The fuels used in ramjet engines are hydrocarbons obtained from petroleum These fuels burn in only a narrow range of fuel air ratios These fuels have flame speeds which are considerably lower than the speed at which the air must pass through the combustion chamber to obtain the high thrust per unit of cross sectional area required for practical applications Flame holders are therefore located in the combustion chamber in the wake of which the air speeds are reduced locally to accommodate the low flame speeds The determination of the configuration and location of these flame holders to provide adequate combustion efficiency without imposing excessive drag on the air is one of the crucial development problems of the ramjet The location and design of the fuel injection nozzles and the design of the control equipment to provide a suitable fuel air ratio in the vicinity of the flame holder for efficient combustion over the range of flight speeds and altitudes desired in the flight program of a given ramjet vehicle are essential to efficient operation New high energy fuels greatly increase the attainable altitude of the ramjet and appreciably reduce the engine length The absence in this propulsion cycle of moving parts after the burner enables the ramjet to burn metal based fuels for greater performance

Take-off. Because the ramjet has low thrust at low flight speeds, another type of engine is required for take off boost In missiles such as the Bomarc, a rocket is used to take off and accelerate the vehicle to a speed at which the ramjet can take over In aircraft where successive take offs and landings are desired, a turbojet engine can be used for this purpose

In general, a supersonic ramjet vehicle must be boosted to supersonic flight speeds before the ramjet engines can provide sufficient thrust for propelling the vehicle By providing diffusers and nozzles in which the configuration and area can be varied, the ramjet can operate efficiently over a wider range of flight speeds and can take over at a lower flight speed This can result in an appreciable saving in the size of the booster rocket

The ramjet depends entirely on pressure recovery due to its forward speed The oblique and normal shock waves at the inlet constitute the first two stages of a compressor deceleration Control of inlet flow by changing the inlet area through axial translation of the nozzle cone or by varying the back pressure through adjustment of the fuel flow can maintain the optimum positions of the shock waves and thereby minimize aerodynamic drag due to spillage or inadequate capture of air

Flight speed. Ramjet engines are usually considered for applications in the range of flight Mach numbers between 2.5 and 8 although 8 is not a theoretical upper limit As the flight Mach number increases above 4, heating of the vehicle by the high air friction becomes progressively a more serious problem, and methods of cooling the structure must be incorporated At flight Mach numbers of 5 and higher, because of the high gas temperatures in the combustion chamber dissociation of the gases occurs, that is, the combustion does not go to completion and only part of the heat is released The remainder of the heat can theoretically be released if combustion continues as the gas expands through the discharge nozzle However, the occurrence of dissociation, the extent of which is influenced by nozzle design may be a basic practical limitation on the flight speeds attainable by ramjets

The working fluid (air) can also be heated by a nuclear reactor, thereby freeing the design from limitations imposed by using the working fluid as oxidizer The air is decelerated passed directly through a nuclear reactor that occupies the usual combustion chamber space and then discharged from a conventional convergent divergent nozzle *See* Combustion wave measurement, Diffuser Nozzle [B P]

Ranales

An order of the plant subclass Dicotyledoneae generally considered to be a primitive genetic group whose earliest members represent the stock from which most of the living angiosperms have descended The order includes 16 families having 328 genera and 5000 species The flower parts are usually numerous distinct and mostly spirally arranged In general the perianth is not differentiated into calyx and corolla The order includes the water lily, buttercup peony, columbine, delphinium, anemone clematis magnolia, tulip tree pawpaw, custard apple moonseed mayapple, barberry nandina, cinnamon camphor tree avocado and many more well known plants Nutmeg is obtained from the seeds of the nutmeg tree *Myristica fragrans* *See* Avocado, Camphor tree, Cinnamon, Magnolia, Nutmeg, Tulip tree, *see also* Dicotyledoneae, Embryophyta, Plant kingdom Tree

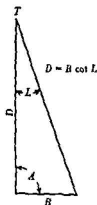


Fig 1 Range triangle

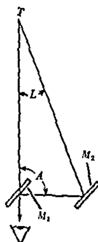


Fig 2 Simple coincidence rangefinder

rangefinder The rangefinder operates as an angle-measuring device for solving the triangle compris ing the rangefinder base length and the line from each window to the target point Rangefinders can be classified in general as being of either the coin cidence or the stereoscopic type

Coincidence rangefinders In these types, one-eyed viewing through a single eyepiece provides the basis for manipulation of the rangefinder ad justment to cause two images or parts of each to match or coincide This type of device is used in its simpler forms in photographic cameras The range triangle for such a rangefinder is shown in Fig 1 where B is the base angle A a right angle D the range and L the convergence angle at the target T The relationship which exists is

$$D = B \cot L$$

The basic optical arrangement is shown in Fig 2, where M_1 and M_2 are a semitransparent mirror and a reflecting mirror respectively When coincidence is obtained that is when the target T is seen in the same apparent position along either path the rangefinder equation is satisfied In many small rangefinders coincidence is achieved by rotating mirror M_2 through a small angle while viewing the images in mirror M_1, other deviating means which do not require rotating of mirror M_2 are also em ployed Figure 3a and b shows the appearance of the images in a superposed field rangefinder If mir ror M_1 is made fully reflecting and arranged to fill only the lower half of the field of view, allowing the eye to view target T directly through the upper half, a split field type of rangefinder is produced; the images appear as shown in Fig. 4a and b

Camera rangefinders. These usually have base lengths varying from less than 1 in. to over 3 in, with magnifications between 0 5 and 2 5. The lower

unity or less is necessary in order to cover the usual 42-52° angle of view.

Military types Coincidence rangefinders of the military type ordinarily employ a penta prism, or its two-mirror equivalent, at each window These prisms, permanently fixed in place, possess the characteristic of reflecting the light rays inward at an angle exactly 90° to the angle of incidence. The light from each prism passes through a telescope objective, and an image of the target formed by each objective is combined and directed by a prism assembly in a single eyepiece for simultaneous viewing Figure 5 shows the geometry of such a rangefinder Here b is the base length of the rangefinder and f is the focal length of the rangefinder's objective lenses Light from an infinitely distant target would arrive along lines TG and $T'H$, while light from a target distance R would arrive along lines TG and TK The objectives, located at G and H and with their axes on lines TG and TH respectively, form images of the target T at J and at K respectively With an infinitely distant target, objective H would form its image at L rather than at K This displacement d, equal to LK, is called the *parallactic* displacement and, since triangles TGH and HLK are similar, is the measure of range The equation is

$$R = \frac{bf}{d}$$

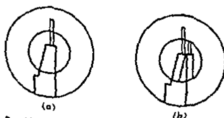


Fig 3 View in superposed field rangefinder (a) Images in coincidence (b) Images unmatched

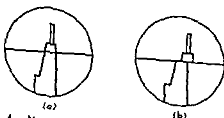


Fig 4 View in split field rangefinder (a) Images matching (b) Images unmatched.



Sheep on a western winter range

The soil is evaluated to determine whether it is normal, eroded, or compacted, and whether its productive potential is increasing or decreasing. It is helpful to know what the forage producing capacity of the soil was originally, what it is at present, and what its possibilities are. The water infiltration rate and the moisture retention capacity of the soil are also measured and, hence, its watershed management condition and its erosion potential determined.

Existing water facilities and possibilities for water development are evaluated to determine whether artificial vegetation rehabilitation is feasible. Accessibility by roads and trails, fire history, and an analysis of fire potential are also determined. In addition, the presence or absence of poisonous plants, predators, and destructive rodents is noted.

Range management plan. The range management plan, developed from the inventory, sets forth the season of year that the range should be grazed, the duration (usually in days) of the grazing period, and the kinds and numbers of heavy livestock (cattle and horses) or lighter animals (sheep and goats) or both to be permitted. The range plan also states the need for and nature of fire protection, the need for additional trails and roads or the abandonment of old ones, the location and types of fences required to direct the control of trailing and herding, the water development needed, any special treatment of soil advisable for eroded or compacted areas, the steps required for control of poisonous plants, and the need for reduction of predatory animals and rodents. The range management plan also details the reseeding needed—the location, type of species, planting methods to be used, use of the vegetation after planting—and considers whether rehabilitation may be expected to take place without artificial seeding under a plan for temporary retirement from use and suitable protection.

Significant advances. Quantitative surveys of range vegetation, a fairly well developed technique in recent years, probably will be improved through use of aerial photos, greater field mobility of survey technicians, and better statistical techniques.

Since about 1935, much has been learned about ... and physiology of range plants (*see* ECOLOGY, APPLIED, PLANT PHYSIOLOGY). This has resulted in the classification of such plants with respect to their physiological condition and the ecological trends indicated, so-called condition and trend classification.

Through quantitative study of animal usage of various plant species, a significant advance has been made in understanding the relation of floristic composition to actual animal carrying capacity of range land (*see* POPULATION DYNAMICS). One method contributing to this knowledge has been the use of the esophageal fistula and fecal collecting bags. Associated with this is an improved and intensive before and after range analysis. The actual nutritional value of the forage species and the usable portions of the plants under various methods of management are determined by laboratory study. This indicates to what extent supplemental feeding of livestock is necessary.

It is now accepted that the management of deer, elk, and antelope herds is also part of range management. Much has been learned concerning the problem of competition of domestic livestock with big game animals, thereby contributing to improved management of range lands for both kinds of animals.

There is growing knowledge of ways to improve

be studied are the adaptability of plants to various locations, methods of planting, physiological adaptiveness of species, time of planting, and the time and type of use after planting. The use of airplanes or helicopters for seeding has not yet solved all range planting problems.

Problems, present and future. There is still question as to whether any significantly large, new range areas can be opened up by an improved or different type of management. The "discovery" for range use of ex timber lands in the coastal plain of the Southeast is an example.

Many problems of multiple use of range land are yet to be resolved. In the western summer range there probably will always be the problem of correlating grazing and timber production. Demands for recreatio[illegible]
coming more [illegible]
lation of the [illegible]
its lands, the demand for water is growing rapidly, making efficient management of range lands imperative for water conservation. *See* FISHERIES CONSERVATION, FOREST CONSERVATION, MINERAL RESOURCES CONSERVATION, SOIL CONSERVATION, WATER CONSERVATION, WILDLIFE CONSERVATION; *see also* LAND USE PLANNING. [L.M.T.]

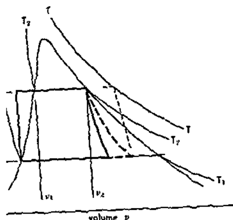
Bibliography. *See* CONSERVATION OF RESOURCES.

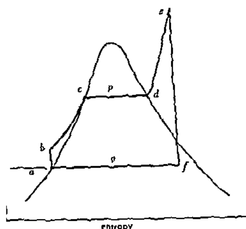
Rangefinder, optical

An optical instrument for measuring distance, usually from its position to a target point. Light from the target enters the optical system through two windows spaced apart, ... ance between the windows being term... len...

... cycle

...dynamic cycle used as an ideal standard ... comparative performance of heat engine ... pump installations operating with a con





...nk ne-cycle diagrams (pressure-volume and temper...ure-entropy) for a steam power plant using super...ated steam

...ensable vapor as the working fluid Applied typi...ally to a steam power plant as illustrated the ...cycle has four phases (1) heat addition *bcde* in a ...boiler at constant pressure p_1 changing water at *b* ...to superheated steam at *e* (2) isentropic expan...sion *ef* in a prime mover from initial pressure p_1 to back pressure p_2 (3) heat rejection *fa* in a con...den er at constant pressure p_2 with wet steam at *f* converted to saturated liquid at *a* and (4) isen...tropic compression *ab* of water in a feed pump from pressure p_2 to pressure p_1

This cycle more closely approximates the opera...tions in a real steam power plant than does the Carnot cycle Between given temperature limits it offers a lower ideal thermal efficiency for the conversion of heat into work than does the Carnot standard Losses from irreversibility in turn make the conversion efficiency in an actual plant less than the Rankine cycle standard See CARNOT CYCLE, REFRIGERATION CYCLE, THERMODYNAMIC CYCLE, VAPOR CYCLE [T B]

Bibliography T Baumeister (ed), *Marks' Mechanical Engineers Handbook* 6th ed 1958, G A Hawkins *Thermodynamics* 2d ed., 1951, J H Keenan, *Thermodynamics*, 1941

Rapakivi granites

A term originally applied to those granites with abundant large ovoid crystals of potash feldspar (orthoclase or microcline) commonly mantled by sodic plagioclase (oligoclase or albite) and em bedded in a matrix of quartz, potash feldspar, plagioclase biotite, and hornblende The ovoid cores may contain grains of the matrix minerals ar ranged in concentric bands The term rapakivi is now commonly applied to any granite with rela tively large mantled crystals of potash feldspar (rapakivi structure) More specifically rapakivi structure embraces an ovoid core mantled by small prisms of oligoclase in more or less radical or tan gential arrangement Best known are the immense bodies of Precambrian rapakivi granite in Finland and Sweden

Many theories have been advanced to explain these puzzling rocks Very likely mantled crystals

...

has resulted from changes in pressure temperature, composition or volatile content, (2) late-magmatic or postmagmatic reconstitution of the rock, or (3) metasomatic replacement of older rocks in some cases in combination with granitization See GRAN ITE GRANITIZATION, MAGMA METASOMATISM [C A CA]

Rape

A member of the cabbage family Rape (*Brassica napus*) does not form a compact head however Rape belongs to the plant order Papaverales Leaves are bluish-green deeply lobed and curled and reach a length of 1½–2 ft Rape has small yel low flowers and tiny black seeds Two types are grown annual for seed and biennial for forage Rape grows well under a wide range of climatic and soil conditions In the United States rape is grown mainly for hog and sheep pasture

Rape for forage is ...

... and Oregon where birdseed varieties are sown Average production of rapeseed in the United States for the two years 1949 and 1954 was 2 220 493 lb valued at $155 659 on the farm Rape for seed is grown more extensively in Europe

Rapeseed is used to some extent for edible pur poses and for soap making It is also a source of oil which is used chiefly in mixtures with min eral oils for lubrication or alone for tempering steel plates The refined oil is known as colza oil

The optical arrangement of a typical military coincidence rangefinder is shown in Fig 6 The rangefinder is adjusted so that if the object being viewed is at infinity both images of the object are seen exactly superimposed If the object is closer than infinity, the image formed by objective G is displaced by a deviation means such as prism D in the manner and amount described by the rangefinder equation, so that superimposition is achieved The amount of such displacement is shown on the scale as the distance reading Other means for obtaining deviation include oppositely rotating wedges (diasporometer), two equal coaxial oppositely arranged prisms with variable spacing, and slidable positive and negative lenses (swing wedge)

Stereoscopic rangefinders. These are entirely different although externally they resemble coincidence rangefinders except for the fact that they possess two eyepieces Both eyes are used with this instrument, rangefinding being accomplished by stereoscopic vision (*see* STEREOSCOPY) It is essentially a large stereobinocular fitted with special reticles which allow a skilled user to superimpose the stereo image formed by the pair of reticles over the images of the target seen in the eyepieces, so that the reticle marks appear to be suspended over the target and at the same apparent distance In the typical stereoscopic rangefinder (Fig 7) the objectives form their images in the plane of their respective reticles A diasporometer in one half of the instrument is used to vary the angle between the beams reaching the eyes from the target until this angle equals that between the beams reaching the eyes from the two reticles, at which point the reticle image will appear to be at the same distance as the target

Rangefinder errors. The accuracy of a rangefinder depends upon the accuracy of the eye in judging the coincidence of two lines as well as upon the characteristics of the instrument and the

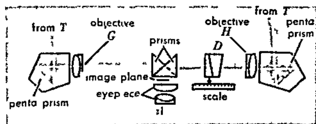


Fig 5 Range triangle for military coincidence rangefinder

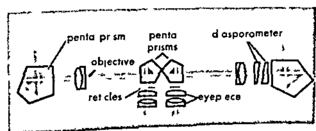


Fig 6 Military coincidence rangefinder

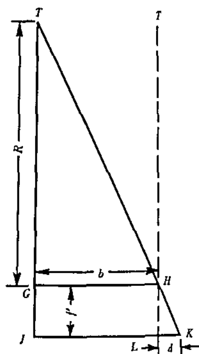


Fig 7 Stereoscopic rangefinder

range distance Rangefinder errors are proportional to the square of the range and inversely proportional to the base length and to the magnification A 1 meter base and a 3 meter base rangefinder of 15 power would have errors of 7 3 and 2 7 yards at 1000 yards range and 470 yards and 170 yards at 8000 yards range respectively

Heightfinders. These are modified forms of rangefinders originally used in the directing of antiaircraft fire A heightfinder depends for its operation upon the knowledge of the range R to the aircraft and its angle of elevation θ The height is

$$H = R \sin \theta$$

In practice one way that the height may be measured is by multiplying range and elevation angle in a gear train computer Another method employs an additional pair of deviation prisms within the rangefinder, geared so as to produce a deviation which varies as $\sin \theta$

Depression rangefinders. These are used in coast defense positions where such instruments can be installed on a cliff top and this elevation is used as the rangefinder base The device measures the depression angle of the target which it converts directly to range after including necessary corrections for tide, curvature of the earth, and refraction of the atmosphere

Stadia methods of rangefinding. These are suitable for some applications If the size of the object is known or can be guessed accurately, its distance can be determined by measuring its size in the eyepiece reticle of a telescopic system *See* LENS, OPTICAL PRISM OPTICAL [E K A]

Bibliography C H v Hofe *Fernoptik*, 1941 D H Jacobs *Fundamentals of Optical Engineering* 1943, A Konig, *Die Fernrohre und Entfernungsmesser*, 1937

mixed crystals or solid solutions. A *single* chemical operation only slightly enriches one rare earth over another, to isolate pure compounds of the individual elements by these methods, the processes have to be repeated many times in order to get [illegible] Historically, the ele-

[illegible] The enormous amount of wo[illegible] permitted the separation of only very small amounts, and the pure rare earths were very costly. For this reason, they gained the reputation of being rare. Fractionation methods are still used commercially to separate crude rare earths and for lanthanum and cerium particularly, since cerium can then be separated from lanthanum by taking advantage of the quadrivalent state of cerium. At present, the other members of the rare earth series are purified by means of ion exchange processes, although if too high purity is not desired, liquid liquid extraction processes can be employed. Some of the rare earths which show anomalous valence are separated from the other rare earths by taking advantage of this valence property.

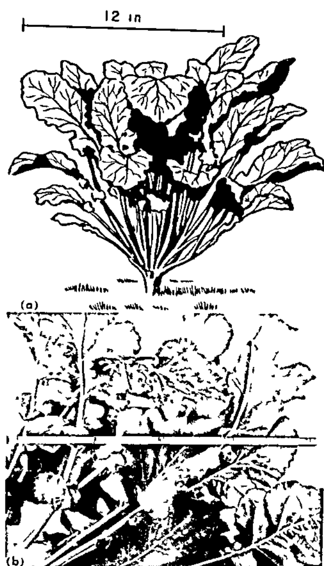
Properties. The elements are metals possessing distinct individual properties which make them potentially valuable for alloying agents. They are usually reduced thermally by treating the anhy-

Table 2. Rare-earth metals

Element	Symbol	Atomic no.	Melting point, °C	Boiling point, °K	Density g/cm³	Atomic volume	Heat of vaporization kcal/mole	Metallic radius, A	Electrical resistivity, ohm-cm, 25°C
Scandium	Sc	21	1538	3000	2.995	15.0	79.0	1.6412	
Yttrium	Y	39	1550	3500	4.472	19.886	93.0	1.802	65×10^{-6}
Lanthanum	La	57	920	3742	6.162	22.5	94.0	1.88	55×10^{-6}
Cerium	Ce	58	804	3741	6.768	20.6	93.0	1.825	74×10^{-6}
Praseodymium	Pr	59	935	3400	6.769	20.9	79.0	1.829	68×10^{-6}
Neodymium	Nd	60	1024	3300	7.007	20.7	69.0	1.823	64×10^{-6}
Promethium	Pm	61							
Samarium	Sm	62	1052	~2200	7.540	20.0	46.0	1.802	88×10^{-6}
Europium	Eu	63	826	1762	5.245	29.3	42.0	1.985	81.3×10^{-6}
Gadolinium	Gd	64	1350	~3000	7.868	20.0	72.0	1.801	144×10^{-6}
Terbium	Tb	65	1360	~3000	8.253	19.3	70.0	1.781	
Dysprosium	Dy	66	1500	~2600	8.565	19.0	67.0	1.773	138×10^{-6}
Holmium	Ho	67	1500	~2600	8.799	18.6	67.0	1.766	87×10^{-6}
Erbium	Er	68	1500–1550	~2900	9.058	18.5	67.0	1.758	185×10^{-6}
Thulium	Tm	69	1500–1550	2000	9.318	18.0	57.0	1.748	79×10^{-6}
Ytterbium	Yb	70	824	~2200	6.959	24.9	38.0	1.9397	27×10^{-6}
Lutetium	Lu	71	1650–1750	~3500	9.819	17.9	90.0	1.733	79×10^{-6}

Symbol	Compressibility cm²/kg	Young's modulus dynes/cm²	Poisson's ratio	Crystal structure at room temperature	Allotropic forms
Sc				Hex., a = 3.3090 A, c = 5.2733 A	
Y	2.09×10^{-6}	6.63×10^{11}	0.265	Hex., a = 3.6474 A, c = 5.7306 A	bcc (above ~1300°C), a = 3.90 A
La	3.2×10^{-6}	3.84×10^{11}	0.288	Hex. a = 3.770 A, c = 12.159 A	fcc (300–868°C), a = 5.301 A; bcc (868–920°C), a = 4.26 A
Ce	4.9×10^{-6}	3.00×10^{11}	0.248	fcc a = 5.1612 A	fcc (below −195°C), a_0 = 4.82 A; Hex., a = 3.68 A, c = 11.92 A; bcc, a = 4.12 A
Pr	3.28×10^{-6}	3.52×10^{11}	0.305	Hex., a = 3.6725 A, c = 11.835 A	bcc (798–955°C), a = 4.13 A
Nd	3.0×10^{-6}	3.79×10^{11}	0.306	Hex. a = 3.6579 A, c = 11.799 A	bcc (868–1024°C), a = 4.13 A
Pm					
Sm	2.56×10^{-6}	3.41×10^{11}	0.352	Rhombohedral, a = 8.996 A, α = 23°13′ (or indexed on hex lattice, a = 3.621 A, c = 26.25 A)	bcc (917–1052°C), a = 4.07 A (very tentative)
Eu	6.94×10^{-6}			bcc a = 4.5820 A	
Gd	2.52×10^{-6}	5.62×10^{11}	0.259	Hex., a = 3.6360 A, c = 5.7826 A	
Tb	2.45×10^{-6}	5.75×10^{11}	0.261	Hex., a = 3.6010 A, c = 5.6936 A	(Transformation noted at 1306°C, from Hex. to unknown form)
Dy	2.39×10^{-6}	6.31×10^{11}	0.243	Hex. a = 3.5903 A, c = 5.6475 A	
Ho	2.14×10^{-6}	6.71×10^{11}	0.255	Hex., a = 3.5773 A, c = 5.6158 A	
Er	2.11×10^{-6}	7.33×10^{11}	0.238	Hex., a = 3.5588 A, c = 5.5874 A	
Tm	2.6×10^{-6}			Hex. a = 3.5375 A, c = 5.5546 A	
Yb	7.12×10^{-6}	1.78×10^{11}	0.284	fcc, a = 5.4862 A	bcc (798–824°C), a = 4.45 A
Lu	2.3×10^{-6}			Hex., a = 3.5031 A, c = 5.5509 A	

Hex. hexagonal, bcc body-centered cubic, fcc face-centered cubic.



Rape, Brassica napus (a) Dwarf Essex, ×1/12 (from L H Bailey, ed, The Standard Cyclopedia of Horticulture, vol 3, Macmillan, 1937) (b) Winter (USDA)

Rapeseed cake, the residue from oil extraction, is valuable for feed and fertilizer *See* CABBAGE, PAPAVERALES, VEGETABLE GROWING [A B B]

Rare-earth elements

The name given to the group of chemical elements starting with element 58 and ending with element 71 The name is a misnomer, since they are neither rare nor earths When the early members of the series were first discovered they were concentrated in the oxide form and since these oxides resemble somewhat the oxides of calcium, magnesium, and aluminum, which were then known as the common earths, they were called the rare earths However cerium is reported to be more abundant in the earth's crust than lead, yttrium more abundant than tin, and even the scarce rare earths, except promethium, are more abundant than the platinum group elements All these elements form trivalent salts which, when dissolved in water, exhibit very similar chemical properties The elements scandium, yttrium, lanthanum and actinium in the IIIa column of the periodic table also show the same similar properties in aqueous solution, and the elements yttrium and lanthanum are always found associated with the rare earths in nature. For this reason, these elements too are very frequently included among the rare earths *See* MONAZITE.

The true rare earths, atomic numbers 58 through 71, arise from the fact that as the atomic number increases in this part of the periodic table the increased charge on the nucleus is compensated by having electrons fill in incomplete inner shells of electrons which play almost no role in the valence forces between atoms This group of elements is sometimes referred to as the lanthanides (*see* LANTHANIDE CONTRACTION)

The elements 90 through 103 also occur in the periodic table at a place where a similar inner shell is being filled In many ways they resemble the lanthanides and are frequently referred to as the actinide rare earths (*see* ACTINIDE ELEMENTS) Both groups are usually displayed at the bottom of the periodic table as an appendage of two rows labeled rare earths *See* PERIODIC TABLE.

Uses. The rare earth metals show great affinity for the nonmetallic impurities such as oxygen, nitrogen, carbon, and hydrogen Therefore, considerable quantities of mixed rare-earth metals are used as "getters" in the metallurgical industry The elements exhibit a very complex spectra when heated and give off an intense white light Mixtures are used in cored carbon arcs in the motion picture industry For commercial uses of the individual rare earth elements, see separate articles on each element

tures in a number of minerals (Table 1) These minerals are usually concentrated from other rock and minerals by mechanical means such as flotation, or magnetic cross belt separation methods The rare earths are leached from the minerals with acid in the case of the phosphate or silicate minerals Some minerals such as the columbotantalates have to be heated with carbon or treated with strong caustic before being leached *See* ORE DRESSING.

Separation. The mixed rare earths can be separated from the acid solutions by means of oxalate precipitation ignition of the oxalate gives a mixed rare earth oxide

The rare earths occur in solution as hydrated trivalent ions whose properties are very much alike and therefore they tend to form precipitates which

Table 1 Some common rare-earth minerals

Mineral	Crystal form	Formula composition*
Monazite	Monoclinic	$CePO_4$ with $Th_3(PO_4)_4$
Xenotime	Tetragonal	YPO_4
Gadolinite	Monoclinic	$2BeO \cdot FeO \cdot Y_2O_3 \cdot 2SiO_2$
Bastnasite	Hexagonal	$CeFCO_3$
Samarskite	Orthorhombic	$3(Fe,Ca,UO_2)_2O \cdot Y_2O_3 \cdot 3(Nb,Ta)_2O_5$
Fergusonite	Tetragonal	$Y_2O_3 \cdot (Nb,Ta)_2O_5$
Euxenite	Orthorhombic	$Y_2(NbO_3)_3 \cdot Y_2(TiO_3)_3 \cdot 1\frac{1}{2}H_2O$
Yttrofluorite	Cubic	$2YF_3 \cdot 3CaF_2$

* Only the most abundant rare earth is listed Ce minerals rich in light members of the rare-earth group Y minerals rich in heavy members of the rare-earth group



Anthracnose d sease on canes (stems) of black erry

r recover Infected plants should be destroyed ıptly and only disease free nursery stock ld be planted Several other species of rust ;ı are found on blackberries and raspberries *irus diseases* These diseases including mo- s, curls, and streaks are the primary cause of of productivity or running-out of cane berries h of these diseases has distinctive symptoms all are systemic and infected plants never re er In addition to being spread by propagation m infected plants the virus diseases usually are nsmitted from diseased to healthy plants by in :ts that feed on the plants Viruses are not known l e spread by pruning or other cultural prac es. Mosaics and leaf curls cause widespread ımage in black and red raspberr es but seldom tack blackl erries The streaks affect black rasp- :rr es and blackberries Measures used for con

or control of insect vectors and use of resistant ır tolerant varieties

Other important diseases Raspberries and blackberries are also susceptible to other impor tant diseases, such as crown gall (*Agrobacterium tumefaciens*) spur blight (*Didymella applanata*), powdery mildew (*Sphaerotheca humuli*) cane blight (*Leptosphaeria coniothyrium*), and fruit rots (*Botrytis* spp) *See* PLANT DISEASE, PLANT VIRUS. [E K V]

Rat

Any of a large number of medium sized rodents usually with long tails found throughout the world The Norway rat *Rattus norvegicus* is the common pest of city and country The name rat in common usage usually refers to this species. The Norway rat is a member of the Old World rats and mice family Muridae but has spread by introduction throughout the world It is an exceedingly destruc tive rodent and is also the common source of bu bonic plague transmitted from infested rats to man by the bite of certain fleas The white rat used in biological research is an albino strain of this species. The roof rat *Rattus rattus* is a related Old World species found along the coasts of the United States especially in the South





(b)

The rat (a) Brown *Rattus norveg cus* length 9 in (b) Black *R rattus* sl ghtly smaller than brown rat (*From P M Duncan ed Cassell s Natural H story Cassell*)

Many New World rodents are called rats These include the white-footed wood rats of the genus *Neotoma* several of which are called pack rats The cotton rats genus *Sigmodon* and the rice rats genus *Oryzomys* are common southern rodents The western long tailed jumping kangaroo rats genus *Dipodomys* are among the most beautiful and interesting mammals. *See* RODENTIA [J D B]

Rat bite fever

There are two diseases transmitted by the bite of a rat, each with distinct etiology and symptomatol ogy One is due to a spiral organism *Spirillum minus* and the other to *Streptobacillus monili*

drous halide with calcium, lithium or other alkali metals and then remelting under vacuum to remove by volatilization the last traces of calcium magnesium and the alkali metals They can also be reduced electrolytically from fused salt baths as is done commercially for cerium and mischmetall (a mixed rare earth metal mainly cerium with small amounts of iron present)

Table 2 gives the properties of the metals The anhydrous solids also show greater difference in properties between the elements than do the hydrated salts

The rare earths form organic salts with certain organic chelate compounds These chelates which have replaced the water around the ions enhance the difference in properties between the individual rare earths and advantage of this is taken in the modern ion exchange methods for separation *See* MAGNETOCHEMISTRY TRANSITION ELEMENTS

[F H SP]

Bibliography D M Yost H Russell and C S Garner *The Rare earth Elements and Their Compounds* 1947

Raspberry

The horticultural name for those species of the genus *Rubus* plant order Rosales in which the fruit when ripe separates thimblelike from the receptacle Raspberry plants are upright shrubs with perennial roots and prickly biennial canes (stems) There are several species both American and European from which the cultivated raspberries have been developed Varieties are grouped as to color of fruit—black red and purple the last being hybrids between the red and black types. Red raspberries with upright canes are propagated by suckers or by root cuttings (Fig. 1) the black varieties with long canes which arch over and touch the ground are propagated by tip layers The hybrid purple varieties are usually propagated by tip layers *See* STEM (BOTANY) Raspberry breeding has been carried on so extensively that all the important red and purple and most of the black varieties are the result of breeding experiments Raspberries are grown extensively in home gardens over most of the United States Leading states in commercial production are Michigan, Oregon New York Washington Ohio Pennsylvania, and Minnesota Annual production usually grosses around $11 000 000 at the farm Although the fruit is sold fresh for dessert purposes it is canned and is made into jelly or jam quick freezing is the most important processing method. See BREEDING (PLANT) FRUIT GROWING (SMALL) ROSALES

[J H CL]

Raspberry and blackberry diseases The diseases of raspberries and blackberries are similar but there is a great difference between the two crops in the amount of damage sustained

Anthracnose This is a disease (*Elsinoe veneta*) occurring on the canes (stems) of all berries most destructive on black raspberries (Fig 2) The disease appears on the lower parts of the stems as round or oval spots with raised purple edges and *sunken grayish centers* Infected canes are brittle, susceptible to winter injury and may dry up and die Anthracnose is controlled by removing immediately after harvest the old canes on which the fungus spends the winter and by spraying with sulfur copper or dithiocarbamate fungicides. A somewhat similar disease leaf and cane spot (*Septoria rubi*) is common on raspberries in eastern North America and on blackberries in the western part This disease is kept under control when the plants are sprayed for anthracnose

Verticillium wilt This disease (*Verticillium albo atrum*) which attacks all cane berries is destructive on black raspberries but seldom causes serious damage to blackberries It is most severe on plants grown in poorly drained soils The disease causes a yellowing wilting and dropping of the leaves that progresses from the ground upward until the cane is defoliated Frequently there is also a bluish discoloration of the canes and eventually the plant dies Only disease free plants should be used and they should be set out in a well drained soil that has been free for at least 4 years from wilt susceptible crops such as potatoes, tomatoes eggplants peppers and strawberries

Orange rust This fungus disease (*Gymnoconia interstitiales*) attacks blackberries and black rasp- The



Fig 1 Red raspberry London variety (USDA)

spicuous mass of orange yellow waxy spores The fungus invades all plant parts and diseased plants

ıbrated ın feet (or meters) per mınute clımb descent ın accord with the altitude pressure on of the standard atmosphere used to calı altimeters [W G B]

ıtes

roup of flightless mostly very large birds, ıerly segregated as a superorder *Palaeognathae* se birds have certain anatomical features ın mon including a sternum without a keel, erfully developed legs an often reduced num of toes lax plumage and large size (except kiwis) The Palaeognathae were supposedly racterized by a "palaeognathous" arrangement he bones of the palate region but S McDowell wed that this palate structure varied from order order The Ratites probably represent several tinct lines of evolution each of which has oduced large flightless birds See AEPYORNITHI RMES APTERYCIFORMES AVES, CASUARIIFOR ES, DINORNITHIFORMES, RHEIFORMES, STRUTHI NIFORMES [K C P]

attlesnake

of the attle ries

lost, so that very large snakes seldom have more than 10–12 rattles at any one time

Rattlesnakes vary greatly ın size The smallest ıs the pigmy rattlesnake *Sistrurus miliarus* of the southeastern and south central United States which is rarely 2 ft long and whose rattle ıs so small that ıt can hardly be heard The related massasauga *S catenatus* found from New York to Mexico ıs only slightly larger The largest are the two species of diamondback rattlesnakes the east ern diamondback *Crotalus adamanteus* and the Texas diamondback *C atrox* Either may reach a length of 90 ın and specimens 6 ft long are com monplace The eastern species is most abundant ın Florida but ranges from North Carolina to Louisi ana in the Coastal Plain *C atrox* occurs from southeastern Missouri south and west to California and northern Mexico ıt ıs most common ın the

of Maine, New Hampshire and northern Vermont Most widely distributed of the eastern rattlesnakes is the timber rattlesnake *C horridus*, which is found in the northeastern part of the United States and westward to the Great Plains and Texas Still common in many places in spite of attempts to wipe it out it is a thoroughly dangerous animal

Except for the smaller species, the bite of any rattlesnake may prove fatal, and even the smaller species or young snakes may inflict a painful bite

Like almost all pit vipers, the young of rattle snakes are born alive Their food consists mainly of warm blooded vertebrates especially rodents, but birds frogs, toads, other snakes, and insects are sometimes eaten See SQUAMATA [J D B]

Rauwolfia

A genus of mostly poisonous, tropical trees and shrubs of the dogbane family (Apocynaceae) Cer tain species are the source of valuable emetics and



Rauwolfia serpentina (L) Benth (From R E Woodson, Jr, H W Youngken E Schlitter, and J A Schneider, *Rauwolfia Botany, Pharmacognosy, Chemistry, and Pharmacology*, Little, Brown, 1957)

cathartics The species *Rauwolfia serpentina*, has received special attention as the source of tranquil izing drugs For centuries, in India, the drug has been used in the treatment of hypertension It came into use in western countries because of its effect in reducing blood pressure Although not a sedative as that term is usually construed, it often has a quieting influence on the patient receiving it. Among the purified alkaloids obtained from *Rau wolfia serpentina*, reserpine (Serpasil) is perhaps the one most used as a tranquilizing agent See GENTIANALES, TRANQUILIZER [P D S]

Raven

A large crow, *Corvus corax*, of the Holarctic Al though the raven is absent from areas of dense populations it is still common in Canada in the

formis *Spirillum minus* is a short, thick, spiral organism 2–5 μ in length, which has three angular curves, polar flagellae, and a rapid darting motion It has not been cultivated in artificial media *See* BACTEROIDACEAE, SPIRILLACEAE

Streptobacillus moniliformis is a gram negative, pleomorphic organism 2–15 μ in length, which grows in chains, interspersed with swollen bodies among the bacillary forms It commonly inhabits the nasopharynx of wild and laboratory animals

Spirillary rat-bite fever. Following a bite the wound heals promptly, but after a period of 5–28 days there is sudden onset of symptoms, including a flare up of the wound site which may subsequently ulcerate, regional lymphangitis and lymphadenitis, chills, relapsing type of fever, macular skin rash, malaise, headache, and often false positive serologic tests for syphilis A leukocytosis is present and the spleen is enlarged Fever may continue for weeks in untreated cases *See* SEROLOGY, SYPHILIS

Definitive diagnosis is by demonstration of organisms under dark field microscopy in exudates from the local lesion and in material aspirated from the enlarged regional lymph node, or by inoculation of these materials intraperitoneally in guinea pigs or mice Penicillin and the tetracyclines are therapeutically effective *See* BACTERIOLOGY, MEDICAL

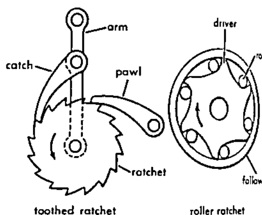
Streptobacillary fever. Most cases of the disease have followed a rat bite, although sporadic cases have been reported in which no direct contact with rats could be established Following a bite, the local wound ordinarily heals, but occasionally an abscess develops The onset of the disease after an incubation period of 1–5 days is abrupt with chills fever, vomiting, headache, severe pains in the back and joints, a maculopapular rash, and leukocytosis Acute arthritis is one of the most prominent and persistent symptoms False positive tests for syphilis are rare Relapses are uncommon

An agglutination titer of 1 80 or higher using the streptobacillus as antigen is regarded as diagnostic, a fourfold or greater rise in titer is especially significant

Penicillin and streptomycin are therapeutically effective *See* PENICILLIN, STREPTOMYCIN [T B T]

Ratchet

A wheel, usually toothed, operating with a catch or a pawl so as to rotate in a single direction, as illustrated A ratchet and pawl mechanism locks a machine such as a hoisting winch so that it does not slip (*see* BRAKE) The locking action may serve to produce rotation in a desired direction and to disengage in the undesired direction as in a drill brace A further adaptation is to drive the catch in a to and fro motion against the ratchet to produce intermittent circular motion The catch or pawl may be of various shapes such as an eccentrically mounted disk or ball bearing Gravity, a spring, or centrifugal force (with the catch mounted internal



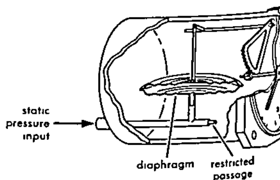
Toothed ratchet is driven by catch when arm moves left, pawl holds ratchet during return stroke of c... In roller ratchet, rollers become wedged betw... driver and follower when driver turns faster than ...lower in direction of arrow

to the ratchet) are commonly used to hold the p... against the ratchet (*see* PAWL) A ratchet and p... provides an arresting action, whereas an esc...ment provides an arresting action followed b... self initiated momentary release (*see* ESC...MENT) In high speed machines, the abrupt ac... of a toothed ratchet produces severe shock In ... situations, a continuously variable yet direction sensitive action is achieved by wedging roller... specially shaped sprags between the input ... output members [F ...

Bibliography P H Black, *Machine Design* ... ed, 1955

Rate-of-climb indicator

An instrument used in aircraft to indicate ver... speed, or rate of climb, and level flight The in...ment contains a small, closed volume which is



Operation of a climb indicator

ternally subjected to static pressure and intern... vented through a capillary tube or its equiva... When the static pressure is changing at a def... rate, as in climb or descent the pressure differ... across the capillary tube is sustained at a v... proportional to rate of climb or descent by th... flow through the tube A metallic corrugated ...phragm forms part of the closed volume and def... proportionally to the pressure difference The

Inductive reactance X_L equals $2\pi fL$ where f is the frequency in cycles per second and L is the self inductance in henries The voltage E across an inductance reaches its peak 90° before the current I reaches its peak, and $I = E/X_L$ amperes Capacitive reactance X_C equals $1/(2\pi fC)$ where C is the capacitance in farads The voltage across a capacitance reaches its peak 90° later than the current reaches its peak and $I = E/X_C$ amperes

Reactances are components of impedance which in general includes resistance R and reactance. Impedance

$$Z = \sqrt{R^2 + (X_L - X_C)^2} \quad \text{ohms}$$

he series *RLC* circuit. In terms of complex ities

$$= R + jX_L - jX_C = R + j(X_L - X_C) \quad \text{ohms}$$

reactances have magnitude and angle $+j$ is $+90°$ for X_L and $-j$ means $-90°$ for X_C ingles by which the voltages across them lead ig the current The resulting phase angle be-n voltage and current is

$$\theta = \arctan\,[(X_L - X_C)/R]$$

current lags, or leads the voltage depending a whether $X_L - X_C$ is positive or negative. ALTERNATING-CURRENT CIRCUIT THEORY

[B L B]

action turbine

prime mover in which fluid enters under pres-e passes through stationary guide vanes where ne of its pressure is converted to velocity and itinues through runner blades where the rest of pressure is expended in producing mechanical ation (see TURBINE) Because pressure drops roughout the machine the proportions of stator d runner must be such that the fluid fills all the ssages In larger sizes the reaction turbine can more efficient than the impulse type (see IMILSE TURBINE) in smaller sizes, the impulse rbine can be the more efficient [F H R]

eactor, electric

device for introducing reactance into a circuit eactors for particular applications usually have pecial designations

Series reactors These reactors (also called urrent limiting) are used in alternating current wer systems for protection against excessively arge currents under short circuit or transient con d ti ns. They are coils of heavy insulated cable ei ther cast in concrete columns or supported in rigid frames and mounted on insulators Magnetic material is absent because high inductance is not needed and current under short circuit would saturate a magnetic core thus reducing reactor effectiveness Series reactors are insulated for greater voltage than other apparatus operated at the same normal line voltage because they reflect incoming voltage surges and increase voltage magnitude In overhead lines exposed to lightning the voltage reflections caused by the reactors are reduced by lightning arrestors or shunting resistors

Shunt reactors These reactors have a relatively

[illegible]

winding Shunt reactors are used to neutra[illegible] charging current of the lines to which they are connected Their insulation and construction is similar to that of power transformers intended for operation on the same kind of power system (see TRANSFORMER) [B L.B , W S P]

Reactor, nuclear

A device containing fissionable material in sufficient quantity and so arranged as to be capable of *maintaining a controlled* self sustaining nuclear fission chain reaction When these conditions are obtained a reactor is said to be in a critical condition For a discussion of critical condition *see* FISSION NUCLEAR, REACTOR PHYSICS The vernacular term pile also identifies nuclear reactors however, its usage is diminishing It originated when the early reactors were constructed by piling blocks of graphite and uranium hence chain reacting pile There are many types of nuclear reactors for various applications *See* REACTOR NUCLEAR (CLASSIFICATION)

Although all reactors produce heat and nuclear radiation they fall in two broad categories, those whose primary purpose is the generation of useful heat or power and those whose *purpose is* the generation of useful *nuclear radiation* In the first category are the various types of power propulsion, and heat generating reactors In the second category are the various types of research, test, irradiation and nuclear fuel production reactors

The principal difference between the two categories of *reactors is the* temperature of the reactor and reactor *coolant In the first* category relatively high temperatures are *required* for efficient conversion of *thermal energy to mechanical energy In* the second category low *temperatures* are preferable because they permit greater ease of utilization of the nuclear radiation produced and incidentally *permit the* use of materials more favorable for the *application See* NUCLEAR AIRCRAFT PROPULSION, NUCLEAR ROCKET REACTOR SHIP PROPULSION

The temperature of the reactor and coolant influences the selection of the coolant the design of the coolant system the mechanical design of the reactor vessel and internals the selection of control systems and reactor safety considerations

Other *considerations* include neutron energy fuel composition fuel distribution moderator size application and cost *See* NUCLEAR POWER

HEAT REMOVAL

The heat generated in a reactor is removed by a primary coolant flowing through the reactor In *most instances the heat is* removed from the primary coolant outside the reactor and the coolant is recirculated Exceptions are the once-through or



The Northern raven, Corvus corax, length to 26½ in (From E L Palmer, Fieldbook of Natural History, McGraw Hill, 1949)

mountainous arid areas of the western United States and along rocky coasts Its American range extends from Alaska and Greenland to Honduras It is present in Michigan and Minnesota in limited numbers, and it is occasionally found as far south on the Atlantic coast as North Carolina The raven is about twice as large as the common crow, but is best distinguished by its hawklike flight It will kill poultry and young lambs, but lives mainly upon carrion, rabbits, and rodents *See* CROW, PASSERIFORMES [J D B]

Raw water

Water obtained from natural sources such as streams, reservoirs, and wells Natural water always contains impurities in the form of suspended or dissolved mineral or organic matter and as dissolved gases acquired from contact with earth and atmos phere Industrial or municipal wastes may also contaminate raw water

In a steam generating unit the contaminates are concentrated by retention and may corrode metals or form heat insulating deposits of sediment or scale on heat transfer surfaces, with resultant overheat ing and possible failure of pressure parts

Principal scale-forming impurities are compounds of calcium and magnesium, or silica Principal cor rosive agents are dissolved oxygen and carbon diox ide In some localities, raw water has a mineral acidity Oil and grease impair wetting and heat removal from the steam generating surfaces and may also form corrosive scale or sludge Certain organic materials or a high concentration of dis solved solids in the boiler water may cause foam ing which contaminates the steam

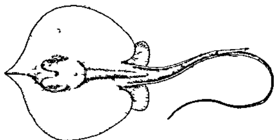
Raw water can be treated to remove objectionable impurities or to convert them to forms that can be tolerated For steam generation, suspended solids are removed by settling (*see* WATER TREATMENT) Scale forming hardness is diminished by chemical treatment to produce insoluble precipitates that are removed by filtration or soluble compounds that do not form scale *See* WATER SOFTENING

Essentially complete purification is achieved by demineralizing treatment or evaporation Deminer alization consists of passing the water through beds of synthetic ion exchange resin particles. Certain of these exchange hydrogen for metallic cations others exchange hydroxyl for sulfate, chloride, or other anions in solution The hydrogen and hydroxyl ions combine to form water The resins may be used in separate or mixed bed arrangements They re quire periodic regeneration by acid and alkaline solutions, respectively, the mixed resins can be separated for such treatment by virtue of their differences in specific gravity

Evaporation requires expenditure of heat for complete vaporization of the water The vapor is subsequently condensed and collected as purified distillate Low pressure steam is used as a heat source, thermal economy is realized through use of multiple effect heat exchange *See* BOILER FEED WATER [F C L]

Ray

Any of several fishes related to the sharks but flat tened dorsoventrally, with their pectoral fins greatly enlarged and broadly joined to the trunk. The large spiracles and eyes are dorsally located whereas the mouth and gill clefts are on the ventral surface



The common sting ray, Dasyatis centroura, length to 3 ft (From E L Palmer, Fieldbook of Natural History McGraw Hill, 1949)

Most rays live in shallow marine waters but a few occur at depths as great as 9000 ft There is one genus of fresh water ray in the Amazon River system Most rays live on the bottom catching their food by ambush tactics but the great manta ray and a few others are free swimming forms that live near the surface in open water

The sting rays have a poison gland opening into the sharp spine on the tail which in some species may inflict a serious, or even fatal sting The tor pedo rays have certain muscle cells modified into electric organs which can release a severe shock. The guitar fish and sawfish are other well known less flattened rays *See* CHONDRICHTHYES, SHARK. [J D B.]

Reactance

The opposition that inductance and capacitance offer to alternating current through the effect of frequency Reactance alters the magnitude of cur rent and also

attractive because of its relatively low melting point (208°F) and high heat transfer coefficient It is also abundant, commercially available in ac ceptable purity and relatively inexpensive It is not particularly corrosive provided low oxygen concentration is maintained Its nuclear properties are fair for thermal reactors and excellent for fast reactors

Sodium presents an activation problem because Na^{24} is formed by the absorption of a neutron and is an energetic gamma emitter with a 15 hour half l fe The containing system requires extensive bio

it decompose and no make-up is required Sodium acts violently with water imposing severe prob- ms in the design of sodium to-water steam boil 's The poor lubricating properties of sodium and s reaction with air further complicate the me- hanical design of sodium cooled reactors The ther alkali metals exhibit similar characteristics nd appear to be less attractive than sodium The utectic alloy of sodium with potassium (NaK) owever has the advantage that it remains liquid it room temperature

Heavy metals have been considered for use as re- actor coolants Uranium is sufficiently soluble in bismuth at high temperatures to permit a liquid fuel system Bismuth also has an extremely small thermal neutron absorption cross section It is a relatively poor heat transfer fluid and in addition the formation of biologically toxic polonium by neutron capture imposes severe leakage restric tions. The high melting point (520°F) of bismuth is also a disadvantage Essentially the same con

a heat transfer fluid it is not a particularly attrac tive reactor coolant It has relatively poor heat transfer and nuclear characteristics and is toxic and expensive

Molten salts Sodium hydroxide (NaOH) has been considered as a reactor coolant but does not appear attractive except perhaps for very special high temperature applications The high melting temperature rather poor nuclear and thermal properties, and corrosiveness of most salts are dif ficult to offset

Fluid flow and hydrodynamics Because heat re- moval must be accomplished as efficiently as possi ble, considerable attention must be given to the fluid flow and hydrodynamic characteristics of the system

The heat capacity and thermal conductivity of the fluid at the temperature of operation have a fundamental effect upon the design of the reactor system The heat capacity determines the mass flow of the coolant required The fluid properties (ther mal conductivity viscosity density and specific heat) are important in determining the surface area required for the fuel to permit transfer of the heat generated at reasonable temperature differ ences This in turn affects the design of the fuel— in particular, the amount and arrangement of the fuel elements These factors combine to establish the pumping characteristics of the system because pressure drop and coolant temperature rise are directly related

Secondary considerations include other physical properties of the coolant, particularly its vapor pressure If the vapor pressure is high at the oper- ating temperature local or bulk boiling of the fluid may occur This in turn must be considered in es tablishing the heat transfer coefficient for the fluid

Because the coolant absorbs and scatters neu trons, variations in coolant density also affect reac tor performance This is particularly significant in reactors where the coolant exists in two phases, for example, the liquid and vapor phases in boiling systems Gases, of course, do not undergo the phase change, nor do liquids operating at temperatures well below their boiling point, however, the fluid density does change with temperature and may have an important effect upon the reactor

Power generation and, therefore, the heat re- moval rate are not uniform throughout the reactor *If the mass flow rate of the coolant is uniform* throughout the reactor, then unequal temperature rise of the coolant results. This becomes particu larly significant in power reactors in which it is de- sired to achieve the highest possible coolant outlet temperature to attain maximum thermal efficiency of the power cycle The performance limit of the coolant is set by the temperature in the hottest re- gion or channel of the reactor Unless the coolant flow rate is adjusted in the other regions of the re actor, the coolant will leave these regions at a lower temperature and thus will reduce the average coolant outlet temperature In high performance power reactors this effect is reduced by orificing the flow in each region of the reactor commensu- rate with its heat generation This involves very careful design and analysis of the system In the *boiling type reactor*, this effect upon temperature does not occur because the exit temperature of the coolant is at the saturation temperature for the sys tem The variation in power generation in the reac tor is reflected by a difference in the amount of steam generated in the various zones

Orificing is also frequently employed in non power systems to achieve a maximum heat re- moval efficiency within the temperature limita tions of the system An example of this type of operation is a research or plutonium production reactor wherein orificing is employed to increase the temperature in the low power generation re gions to minimize the coolant flow through the re- actor All of the coolant does essentially the same amount of work (removal of heat) with the mini mum total coolant flow rate This minimizes the pumping power required and results in maximum efficiency of heat removal

In very high performance reactors, the flow rate and consequent pressure drop of the coolant are

single pass systems that employ water or air coolant The circulating fuel type reactors include fuel as a solution or suspension in the coolant These are of the recirculating type but impose the additional considerations of fuel circulation external to the reactor

Heat is not generated uniformly in a reactor The heat flux decreases axially and radially from a peak at the center of the reactor, or near the center if the reactor is not symmetrical in configuration In addition, local perturbations in heat generation can occur because of inhomogeneities in the reactor structure These variations impose special considerations in the design of reactor cooling systems, including the need for establishing variations in coolant flow rate through the reactor to achieve uniform temperature rise in the coolant; avoiding local hot spot conditions, and avoiding local thermal stresses and distortions in the structural members of the reactor

Nuclear reactors have the unique thermal characteristic that heat generation continues after shutdown because of fission and radioactive decay of fission products Significant fission heat generation occurs for only a few seconds after shutdown *Radioactive-decay heating varies with the decay char*acteristics of the fission products

Accurate analysis of fission heat generation as a function of time immediately after reactor shutdown requires detailed knowledge of the speed and reactivity worth of the control rods The longer term fission product decay heating depends upon prior reactor operation Typical values of the total heat generation after shutdown (as per cent of operating power) are *10–20%* after *1* sec, *5–10%* after 10 sec, approximately 2% after 10 min, 1 5% after *1* hour and *0 7%* after 1 day

Reactor coolants. Coolants are selected for specific applications on the basis of their heat transfer capability, physical properties and nuclear properties There is no one fluid that has optimum properties for all reactors Coolants that have received most attention are light water (H_2O), heavy water (D_2O), air carbon dioxide (CO_2), helium (*He*), diphenyl ($(C_6H_5)_2$), sodium (Na), and sodium potassium alloy (NaK) Other coolants that have been considered but have not found general acceptance include bismuth (Bi), lead bismuth (Pb Bi), mercury (Hg), and various molten salts

Water Water has many desirable characteristics and was employed as the coolant in the first production reactors Both light and heavy water are excellent *neutron moderators*, whereas heavy water (deuterium oxide) has a neutron absorption cross section approximately 1/600 that for light water

Water is a reasonably attractive heat transfer fluid, its technology is well developed, it is abundant and economical There is no serious neutron activation problem with pure water; N^{16}, formed by the (*n p*) reaction with O^{16} (absorption of a neutron *followed* by emission of a proton), is the major source of activity, but its 7.5 sec half life minimizes this problem The most serious limitation of water as a coolant for power reactors is its high vapor pressure A coolant temperature of 550°F requires a system pressure of approximately 1500 psi This temperature is far below modern power station practice where temperatures in excess of 1000°F have become common Lower thermal efficiencies result from lower temperatures. The high pressure necessary for water-cooled power reactors imposes severe design problems, which will be discussed later in this article.

Gases Gases are inherently poor heat transfer fluids as compared with liquids because of their low density This situation can be improved by increasing the gas pressure, however, this introduces other problems Helium is the most attractive gas (it is chemically inert and has good thermodynamic and nuclear properties) It is expensive and not readily available outside the United States. It is difficult to contain, and leakage contributes significantly to the cost of power produced Gases are capable of operation at extremely high temperature and they are being considered for special process applications and direct cycle gas turbine applications

Organic coolants Diphenyl and terphenyl possess good neutron moderating properties and have lower vapor pressures than water Organic coolants are noncorrosive and relatively inexpensive Their major disadvantage is dissociation or decomposition under irradiation The decomposition products can be removed by distillation, however the cost of replacement of coolant contributes to the cost of power produced

Liquid metals The alkali metals, in particular have excellent heat transfer properties and extremely low vapor pressures at *temperatures* of interest for power generation Sodium is the most

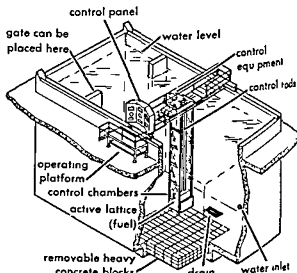


Fig 1 Bulk shielding reactor This light water moderated pool reactor is of the heterogeneous enriched fuel type and provides high thermal neutron flux, ready accessibility, and versatility (*From U.S Atomic Energy Commission, Research Reactors, McGraw-Hill, 1955*)

cost than a more conventio ...

In addition to the variety of special pumps developed for reactor coolant systems there is a variety of piping system components and heat exchange components. As in all flow systems flow regulating devices such as valves are required as well as flow instrumentation to measure and thereby control the systems Here again leak tightness has necessitated the development of special valves with metallic bellows around the valve stem o ensure system integrity Measurement of flow nd pressure has also required the development of ensing instrumentation that is reliable and leak ight

Many of these developments have borrowed from other technologies because toxic or inflammable luids are frequently pumped in other applications In many cases however special equipment has been developed specifically to meet the re

principle of a moving conductor in a magnetic field is employed by placing a magnet around the pipe and measuring the voltage generated by the moving conductor (coolant) in terms of flow rate Temperature compensation is required and calibration is critical

Although the development of nuclear power reactors has introduced many new technologies it has not yet displaced the conventional steam cycle as the most efficient method for converting thermal energy to mechanical energy Steam is generated either directly in the reactor (direct cycle boiling reactor) or in auxiliary steam generation equipment in which steam is generated by transfer of heat to water from the reactor coolant. These steam generators have required very special design particularly when dissimilar fluids are involved Typical of these problems are the sodium to-water steam generators in which absolute integrity is essential because of the violent chemical reaction between sodium and water

CORE DESIGN AND MATERIALS

The reactor core consists of the fuel elements (usually rods or plates) supported by a grid type structure inside a vessel or tank and neutron moderators

The primary function of the vessel is to contain the coolant Its design and materials are determined by such factors as the nature of the coolant (corrosive properties) operating conditions (temperature and pressure) and quantity and configuration of fuel To complicate vessel design further the vessel is pierced by devices for controlling reactor operation for loading and unloading the fuel and for coolant entrance and exit.

Design must also take account of thermal stresses caused by temperature differences in the system Another problem is radioactivity induced in core materials because of neutron absorption during reactor operation This precludes normal maintenance of the equipment and in some areas makes repairs virtually impossible For this reason an exceptionally high degree of integrity is demanded of this equipment Some reactors have been designed to permit removal of the internals from the vessel this is difficult however and tends to complicate the design of the system

Structural materials Structural materials employed in reactor systems must possess suitable nuclear and physical properties and must be compatible with the reactor coolant under the conditions of operation Additional requirements such as fuel cladding are imposed by certain applica t n Some requirements are especially severe be-

base material

The most common structural materials employed in reactor systems are aluminum steel and zirconium alloys Aluminum and zirconium alloys have favorable nuclear properties, whereas the steel alloys have favorable physical properties Aluminum is widely used in low temperature reactors zirco-

absorption is critical

The 18–8 series stainless steels have been used for structural members in both water cooled reactors and sodium cooled reactors because of their corrosion resistance and favorable physical properties at high temperatures Type 304 and type 347 stainless steel have been used most extensively because of their weldability machinability and physical properties To reduce cost heavy walled pressure vessels are normally fabricated from carbon steels and clad on the internal surfaces with a thin layer of stainless steel to provide the necessary corrosion resistance Vessels of this type have been constructed up to 9 ft in diameter and with total wall thicknesses up to 8 in

Although pressure vessels have been constructed for other industries to meet even more severe service requirements the complex requirements for reactors have introduced new design and fabrication problems Of particular importance is the dimensional precision required and the special nozzles and other appurtenances required

The large gas-cooled power reactors require pressure vessels that preclude transportation as a single unit This has necessitated the development of field fabrication techniques including field welding of wall sections up to 3 in thick and subsequent stress relieving of the welded structure. Many unique problems have required the development of new techniques in heavy-steel fabrication

sufficiently high to create mechanical problems in the system It is not uncommon for the pressure drop through the fuel assemblies to exceed the weight of the fuel elements in the reactor with a resulting hydraulic lifting force on the fuel elements Often this requires a design arrangement to hold the fuel elements down Although this can be overcome by employing downward flow through the system, this is often undesirable because of shutdown cooling considerations It is very desirable in most systems to accomplish shutdown cooling by natural convection circulation of the coolant If downflow is employed for forced circulation, then shutdown cooling by natural convection circulation requires a flow reversal, which can introduce new problems

Thermal stress considerations. The temperature of the reactor coolant increases as it circulates through the reactor This increase in temperature is constant at steady state conditions Fluctuations in power level or in coolant flow rate result in variations in the temperature rise These are reflected as temperature changes in the coolant exit temperature, which in turn result in temperature changes in the coolant system

A reactor is capable of very rapid changes in power level, particularly, reduction in power level Reactors are equipped with mechanisms (reactor scram systems) to ensure rapid shutdown of the system in the event of an operational abnormality

Therefore, reactor coolant systems must be designed to accommodate the temperature transients that can occur because of rapid power changes In addition, they must be designed to accommodate temperature transients that might occur as a result of a coolant system malfunction such as pump stoppage The consequent temperature stresses induced in the various parts of the system are superimposed upon the thermal stresses that exist under normal steady state operations

In very high performance systems it is not uncommon for the thermal stresses alone to approach the allowable stresses in the materials of construction In these cases careful attention must be given to the transient stresses, and thermal shielding is

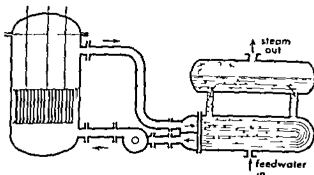


Fig 2 Pressurized water reactor Pump circulates water between reactor tank and boiler Coolant also acts as moderator in natural or slightly enriched core (From L N Rowley and B G A Skrotzki, eds Nuclear Energy Today, Power spec rept, Dec, 1955)

commonly employed in critical sections of the system Normally, this consists of a thermal barrier, which, by virtue of its heat capacity and resistance to heat transfer, delays the transfer of heat, thereby reducing the rate of change of temperature and protecting critical system components from thermal stresses

Thermal stresses are also important in the design of reactor fuel elements Metals are frequently required that possess dissimilar thermal expansion coefficients Heating of such systems gives rise to distortions, which in turn can result in flow restrictions in coolant passages Very careful analysis and experimental verification are often required to avoid such circumstances

Reactor-coolant-system components The development of reactor systems has necessitated concurrent development of special components for reactor coolant systems These have been required even for systems employing conventional coolants such as water or air

Because of the hazard of radioactivity, leak tight systems and components are a prerequisite to safe, reliable operation and maintenance Special problems are introduced by many of the fluids employed as reactor coolants Heavy water systems have special out leakage considerations because of the very high cost of heavy water In leakage should also be prevented because air contains water vapor which tends to dilute the heavy water, necessitating expensive purification

The more exotic coolants developed for reactor applications have required much more extensive component development The alkali metals (sodium, NaK and potassium) are chemically very active and are extremely poor lubricants Centrifugal pumps must be specially designed, employing unique bearings and seals Conventional bearings are completely unsatisfactory in a sodium environment However hydraulic bearings that are quite successful have been developed Satisfactory seals for sodium under pressure have not yet been developed, however it has been possible to design sump type pumps with a free liquid surface within the pump and an inert gas blanket above this surface The seals are located in the inert gas region and only the gas must be sealed Even here considerable difficulty has been encountered because of the sodium vapor contained in the gas The early pumps were designed with the entire motor enclosure contained within a gas tight envelope This procedure is still being employed in some reactor coolant systems, however application becomes more difficult as larger pumping systems are required Because liquid metals are excellent electrical conductors electromagnetic type pumps have been developed These pumps are completely sealed contain no moving parts and derive their pumping action from electromagnetic forces imposed directly on the fluid *See* ELECTROMAGNETIC PUMPS, SODIUM

Totally enclosed pumps have also been developed for water cooled reactors They are com-

a pletely sealed units and can be welded directly into si the coolant piping systems The water *is circulated* x through the electric motor to provide the necessary ly *cooling* Although they have been quite successful ...derably higher

pumps de
...ere is a va
...ety of piping system components and heat ex hange components As in all flow systems ow regulating devices such as valves are required s well as *flow instrumentation* to measure and hereby control the systems Here again *leak tight* ess has necessitated the development of special alves with metallic bellows around the valve stem o ensure system integrity Measurement of flow and pressure has also required the development of sensing instrumentation that is reliable and leak tight

Many of these developments *have borrowed from* other technologies because toxic or inflammable fluids are frequently pumped in other applica tions In many cases however special equipment has been developed specifically to meet the re*quirements* of the reactor systems An example of

is employed by placing a magnet around the pipe and measuring the voltage generated by the moving conductor (coolant) in terms of flow rate Temper ature compensation is required and calibration is *critical*

Although the development of nuclear power re actors has introduced many new technologies it has not yet displaced the conventional steam cycle as the most efficient method for converting thermal energy to mechanical energy Steam is generated either directly in the reactor (direct cycle boiling reactor) or in auxiliary steam generation equip*ment in which steam is generated by transfer of* heat to water from the reactor coolant These steam generators have required very special design par ticularly when dissimilar fluids are involved Typi cal of these problems are the sodium to-water steam generators in which absolute integrity is es sential because of the violent chemical reaction be tween sodium and water

CORE DESIGN AND MATERIALS

The reactor core consists of the fuel elements (usually rods or plates) supported by a grid type structure inside a vessel or tank and neutron mod erat rs

The primary function of the vessel is to contain the coolant Its design and materials are deter mined by such factors as *the nature of the* coolant (corrosive properties) operating conditions (tem perature and pressure) and quantity and configu ration of *fuel* To complicate vessel design further the vessel is pierced by devices for controlling re actor operation, for loading and unloading the fuel and for coolant entrance and exit

Design must also take account of thermal stresses caused by temperature differences in the system Another problem is radioactivity induced in core materials because of neutron absorption during *reactor operation* This precludes normal maintenance of the equipment and in some areas makes repairs virtually impossible For this reason an exceptionally high degree of integrity is de manded of this equipment Some reactors have been *designed to permit* removal of the internals from the vessel this is difficult however and tends to complicate the design of the system

Structural materials Structural materials em ployed in reactor systems must possess suitable *nuclear* and physical properties and must be com patible with the reactor coolant under the condi tions of operation Additional requirements such as fuel cladding are imposed by certain applica tions Some requirements are especially severe be cause of secondary effects for example corrosion limits may be established by the rate of deposition of coolant entrained corrosion products on critical surfaces rather than by the rate of corrosion of the base material

The most common structural materials employed in reactor systems are aluminum steel and zirconium alloys Aluminum and zirconium alloys have *favorable nuclear properties whereas the steel al* loys have favorable physical properties Aluminum is widely used in low temperature reactors zirconium and steel are used in high temperature reac tors Zirconium is relatively expensive and its use *is therefore confined to applications where neutron* absorption is critical

The 18–8 series stainless steels have been used for structural members in both water cooled reac tors and sodium cooled reactors because of their corrosion resistance and favorable physical *prop* erties at high temperatures Type 304 and type 347 stainless steel have been used most extensively be *cause of their* weldability machinability and phys ical properties To reduce cost heavy walled pressure vessels are normally fabricated *from carbon* steels and clad on the internal surfaces with a thin layer of stainless steel to provide the necessary corrosion resistance Vessels *of this type* have been constructed up to 9 ft in diameter and with total wall thicknesses up to 8 in

Although pressure vessels have been constructed for other industries to meet even more severe serv ice requirements the complex *requirements for re* actors have introduced new design and fabrication problems Of particular importance is the dimen sional *precision* required and the special nozzles and other appurtenances required

The large gas cooled power reactors require pressure vessels that preclude transportation as a single unit This has necessitated the development of field fabrication techniques *including* field welding of wall sections up to 3 in thick and subse quent stress relieving of the welded structure Many *unique problems* have required the development of new techniques in heavy steel fabrication

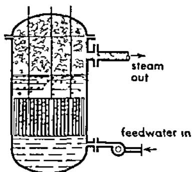


Fig 3 Boiling water reactor Steam is produced in water moderated enriched core (*From L N Rowley and B G A Skrotzki, eds, Nuclear Energy Today, Power spec rept, Dec, 1955*)

Research reactors operating at low temperatures and pressures introduce special experimental considerations The primary objective is to provide the maximum volume of unperturbed neutron flux for experimentation It is desirable, therefore, to extend the experimental irradiation facilities beyond the vessel wall This has introduced the need for vessels constructed of materials having a low cross section for neutron capture Relatively large aluminum reactor vessels with wall sections as thin as practicable have been manufactured for research reactors Special problems with respect to dimensional stability have necessitated unique supporting structures The vessel design is complicated further by the variety of openings that must be provided to accommodate experimental apparatus It is highly desirable to provide access to the reactor proper for experiments and in many cases, to have apparatus installed in so called through holes that penetrate the vessel from side to side

In some instances, stainless steel vessels have been employed for research and test reactors at the sacrifice of some experimental flexibility The experimental irradiations are performed within the reactor vessel and limited use is made of the space external to the reactor vessel

A special problem is introduced by research reactors employing heavy water as a moderator and light water as a coolant A calandria type design has been employed, consisting of an all aluminum multitube container for the heavy water with additional aluminum tubes connected to separate coolant headers for circulation of the light water coolant This arrangement introduces the special problems associated with the multitudinous welds to contain a system within a system each being tight with respect to leakage to the atmosphere and to the other system

Fuel cladding. Heterogeneous reactors maintain a separation of fuel and coolant by cladding the fuel The cladding material must be compatible with both the fuel and the coolant

The structural materials must also have favorable nuclear properties The neutron capture cross section is most significant because the parasitic absorption of neutrons by these materials reduces the efficiency of the nuclear fission process Aluminum is a very desirable material in this respect however, its physical strength and corrosion resistance in water decrease very rapidly above about 300 F Some improvement in its corrosion resistance has been obtained by the addition of small amounts of nickel to the alloy, however, it is quite improbable that these qualities can be improved for service beyond 500°F

Zirconium has very favorable neutron properties, and in addition can be made reasonably corrosion resistant in high temperature water It has found extensive use for water cooled power reactors, although it is expensive and difficult to fabricate. The technology of zirconium and zirconium base alloys has advanced tremendously under the impetus of the various reactor development programs.

Stainless steel has been employed as a fuel cladding However, because of its relatively large neutron capture cross section, it must be used in very thin sections

The fuel and the cladding must be thermally bonded to achieve maximum heat transfer Zirconium can be metallurgically bonded to uranium metal to form an integral structure Aluminum can be bonded to a uranium aluminum mixture in a technique employed quite extensively in the fabrication of fuel elements for research reactors. Other combinations of fuel and clad that do not permit an effective metallurgical bond have achieved efficient heat transfer through a bonding agent such as lead or sodium *See* NUCLEAR FUELS

Fuel materials. Uranium metal is susceptible to irradiation damage which limits its operating life in a reactor The life expectancy can be improved somewhat ... derably more by ... or molyb... *radiation damage resistance,* and in addition ...

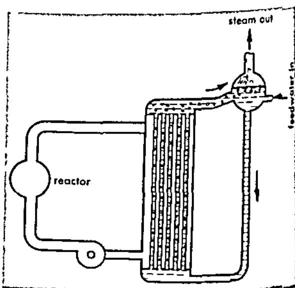


Fig 4 Homogeneous reactor Fluid bearing fuel is circulated between reactor and boiler Critical mass in reactor heats fluid (*From L N Rowley and B G A Skrotzki eds Nuclear Energy Today, Power spec rept, Dec, 1955*)

rrosion resistant in oxidizing media Ceramics ch as uranium oxide have a very low thermal con ıctivity and lower density than metals which is sadvantageous in certain applications

Uranium metal can be fabricated by relatively ell established techniques provided proper care taken to prevent oxidation The metal is melted vacuum furnaces and can be cast by gravity or ıjection. Ingots can be rolled or extruded and elatively complicated shapes can be fabricated lost commonly fuel elements are in the shape of ods or plates and are fabricated by casting roll ng, or extrusion Uranium oxide is fabricated by ompacting and sintering frequently into pellets hat are then sealed in metal cylinders

Plutonium is a very difficult material to handle ... ery undesirable physical properties for

widely used for nuclear reactors because it is potentially more abundant than uranium 235

Uranium 233 also imposes special handling problems because of biological toxicity but it does not introduce new metallurgical problems Thorium the fertile material source for uranium 233

Moderators Reactors designed to operate with thermal (slow) neutrons require a material called a moderator to slow down the high energy fission neutrons to the thermal range

The best moderating materials are among the low atomic weight elements such as the hydrogen isotopes beryllium and carbon They have very low neutron capture cross sections and because of their lightness reduce the kinetic energy of the neutrons rapidly in the fewest number of collisions See RADIATION DAMAGE (INANIMATE MATERIALS)

Light and heavy water are convenient to use because they can serve as coolants as well as moderators. Liquid hydrocarbons such as terphenyl are also being considered although such organic materials are subject to radiation damage

Among the solid moderators are carbon (graphite) and beryllium metal or beryllium oxide The graphite used in reactors is more highly purified than that used in nonreactor applications to minimize neutron capture See BERYLLIUM GRAPHITE.

CONTROL AND INSTRUMENTATION

The control of reactors requires the measurement and adjustment of the critical condition. A reactor is critical when the rate of production of neutrons equals the rate of consumption in the system. The neutrons are produced by the fission process and are consumed in a variety of ways, including absorption to cause fission, nonfission capture in fissionable materials, capture in fertile materials capture in structure or coolant, and leakage from the reactor A reactor is subcritical (power level decreasing) if the number of neutrons produced is less than the number consumed The reactor is supercritical (power level increasing) if the number of neutrons produced exceeds the number consumed

Reactors are controlled by adjusting the balance between neutron production and neutron consumption Normally, neutron consumption is controlled by varying the absorption or leakage of neutrons, however, the neutron generation rate can be controlled by varying the amount of fissionable material in the system

It is essential to orderly control and management of a reactor that the neutron density be sufficiently high to permit reliable measurement A source of neutrons is essential therefore, to the control and instrumentation of reactor systems Neutrons are obtained from the photo-neutron effect in materials such as beryllium Neutron sources consist of a photon (γ ray) source and beryllium, such as antimony beryllium or radium beryllium Antimony sources are usually preferred because the antimony is activated by the reactor neutrons each time the reactor operates, whereas the strength of radium sources decreases with time For discussion of the principles of reactor control, see REACTOR PHYSICS

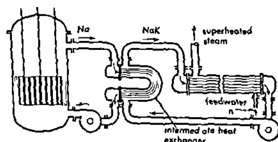


Fig 5 Sodium-graphite reactor Two liquid metal coolant circuits with intermediate heat exchanger avoid making steam radioactive in once-through boiler (From L N Rowley and B G A Skrotzki, eds, Nuclear Energy Today Power spec rept, Dec, 1955)

Control drives and systems The reactor control system requires the movement of neutron absorbing rods (control rods) in the reactor under very exacting conditions They must be arranged to increase reactivity (increase neutron population) slowly and under absolute control They must be capable of reducing reactivity both rapidly and slowly

Normal operation of the control drives can be accomplished manually by the reactor operator or by automatic control systems. Reactor scram (very rapid reactor shutdown) can be initiated automatically by one or more system scram safety signals or manually by depressing a scram button convenient to the operator in the control room

Control drives are normally electromechanical devices that impart linear or swinging motion to the control rods They are usually equipped with a

relatively slow speed reversible drive system for normal operational control Scram is usually effected by a high speed overriding drive accompanied by unlatching or disconnecting the main drive system To enhance reliability of the scram system, its operation is usually initiated by de-energizing appropriate electrical circuits This also automatically produces reactor scram in the event of a system power failure Hydraulic or pneumatic drive systems have also been developed, as well as a variety of electromechanical systems

In addition to the actuating motions required, control rod drive systems must also provide accurate indication of the rod positions at all times Various types of selsyn drive are employed as position indicators as well as arrangements of switches and lighting systems. It is not uncommon to provide control rod position indication accurate to a few thousandths of 1 in

Reactor instrumentation. Reactor control requires measurement of the reactor condition Neutron sensitive ion chambers are used to measure neutron flux These neutron detectors are normally located near the extremity of the reactor and measure an average flux that is proportional to the average neutron density in the reactor The chamber current is calibrated against a thermal power measurement and then applied over a wide range of reactor power level The neutron sensitive detectors must respond to the lowest neutron flux in the system produced by the neutron source

Normally, many channels of instrumentation are required to cover the entire operating range Several channels are required for low level operation, beginning at the source level, whereas others are required for the intermediate and high power level ranges Ten channels of detectors are not uncommon in reactor systems The total range to be covered is in the range of 7–10 decades of power level

The chamber current can be employed as a signal, suitably amplified to operate automatic control system devices as well as to actuate reactor scram In addition to absolute power level, rate of change of power level is also an important measurement which is recorded and employed to actuate various alarm and trip circuits The normal range for the current ion chambers is approximately 10^{-14}–10^{-4} amp This current is suitably amplified in logarithmic and period amplifiers, and can be measured directly with a galvanometer

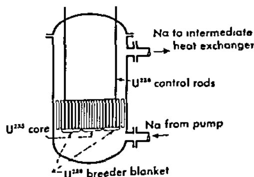


Fig 6 Fast breeder reactor This type includes U^{235} core, U^{238} blanket, liquid metal coolant (From L N Rowley and B G A Skrotzki, eds, Nuclear Energy Today, Power spec rept, Dec, 1955)

REACTOR SAFETY

Special consideration must be given to reactor safety because reactors contain large quantities of lethal substances Fission products and some of the heavy elements produced are extremely toxic.

Safety considerations can be divided into two general categories: (1) hazard to the plant and its occupants, and (2) hazard to the population beyond the environs of the plant The first category is treated as a plant safety problem somewhat analogous to other critical industrial operations, and radiation monitoring is employed extensively for the protection of operating personnel The second category receives intensive study Of particular significance is the evaluation of possible energy release sufficient to disperse the reactor products to the environment Consideration is given to the possibility of chemical energy release as well as of nuclear energy release In some instances, the stored energy in the system is also a significant factor The safety evaluation is normally directed toward the assurance that the maximum credible accident will not result in the release of significant radioactivity to the environment

One of the most important considerations in evaluating reactor safety is the power characteristics of the reactor, in particular, any circumstances that would tend to make the reactor autocatalytic (causing an increase in power tending to produce a further increase in power) Most reactors can be designed to possess a negative temperature and power coefficient, that is, a reactivity decrease would accompany a temperature or power increase.

An increase in temperature causes thermal expansion of the system and a reduction in density of the various components These can be utilized to effect a reduction in reactivity of the reactor Where such reduction cannot be accomplished (or where its accomplishment is doubtful), it is important that the positive coefficient be small and slow in response This too can normally be accomplished by proper design

The safety or scram system of the reactor must operate rapidly and must be initiated early enough in a power rise to prevent reactor damage. In addition different signals should initiate reactor shutdown (abnormal neutron flux, temperature, and coolant flow) and multiple shutdown devices independently operated are required to provide reliability

Careful consideration must be given to the possible minor malfunctions that could conceivably initiate a succession of major difficulties For example a small rupture in the coolant system could result in the loss of coolant which in turn, could promote the melting of the fuel which in turn, could result in a supercritical assembly or a chemical reaction

Special attention must be given to the shutdown ... actor to remove fission product

In some power reactors ...

under pressure which ...
... energy in liquid sodium which can react

...ch in the interests of economy ... be located in close proximity to populated ...as.

This problem has been resolved by containing ...ctor systems in gas-tight structures The re...rements for the containment system are estab...hed by the characteristics of the reactor system ...here a potentially large energy release can ac...mpany or initiate the release of radioactivity the ...ntainment system must be capable of withstand...g the resultant pressure and temperature without ...ss of integrity Spherical and cylindrical contain...ent vessels up to 225 ft in diameter and capable ...f withstanding internal pressure in excess of 30 ...si have been constructed to house various reactor ...ystems Special air locks are provided to permit ...ntrance and exit of personnel and equipment with...ut violating the gas tight requirements of the ...tructure. Special provisions have been made in the ...ASME Boiler Code to establish requirements for ...hese structures See RADIATION SHIELDING

POWER GENERATION SYSTEMS

The thermal energy produced in a nuclear reactor must be converted to mechanical energy for direct utilization or for the generation of electricity This is ordinarily achieved with a steam cycle The steam is generated in a nuclear power plant at a lower temperature than in conventional fossil fueled plants

In the case of the direct cycle boiling reactor the coolant is the working fluid in the power cycle In all other reactor types the coolant is an intermediate fluid from which the heat must be transferred to the working fluid of the power cycle

The gas-cooled reactor is most closely analogous to the conventional fossil fueled plant in that gas at high temperature transmits the heat to the steam generation equipment. The gas coolant from the reactor however is at a much lower temperature than the combustion gases in a furnace because it must be at a lower temperature than the fuel to effect heat transfer

The reactor coolant is also radioactive because of the absorption of neutrons in its passage through the reactor The intensity of the radioactivity and thus the magnitude of the problem is dependent on the nuclear characteristics of the coolant. Activation of water is relatively low whereas the activation of sodium is extremely high In a direct cycle boiling reactor, contamination of the entire turbine and condensing system is possible if activated impurities are permitted to deposit in these units The indirect cycle reactors confine possible contamination to the closed primary coolant system Beyond the steam generation equipment the power generation system is very similar to conventional power systems. [L.J.K.]

Bibliography C F Bonilla (ed), *Nuclear En...* ...d), *Nuclear* ...sstone, *Prin...* ..., 1955, D B Hoisington *Nucleonics Fundamentals* 1959, R L. Murray, *Introduction to Nuclear Engineering* 1954, R Stephenson, *Introduction to Nuclear Engineering*, 2d ed , 1958

Reactor, nuclear (classification)

Nuclear reactors are used in a variety of ways as sources for nuclear radiations and energy During the years following the first reactor demonstration in 1942 intense and world wide reactor development programs have been launched New developments in materials for fuel moderator, reflector, coolant, control, and shielding have expanded greatly the number of promising types and combinations for reactors Thus extreme reactor diversification has evolved with reactor dimensions varying from football size to house-size and rates for energy release from a fraction of a watt to hundreds of megawatts The developments are still too new to single out optimum reactor types and designs as power producers A convenient approach to obtaining a perspective of the development of nuclear fission reactors is to classify reactors by basic design selections and by applications

CLASSIFICATION BY DESIGN

Design parameters used to classify reactors are neutron energy fuel composition fuel distribution and coolant For more detailed description of reac...

...TION DAMAGE (INANIMATE MATERIALS), RADIATION SHIELDING RADIOISOTOPE PRODUCTION, REACTOR NUCLEAR, REACTOR PHYSICS, REACTOR SHIP PROPULSION

Neutron energy and moderator. A fission reactor is an assembly of materials so arranged that the neutron chain reaction is self sustaining and controllable Neutrons are required to produce fission and in turn neutrons are released by the fissioning of the fuel The fission neutrons released at very high energies (1-2 Mev) and correspondingly high speeds (one-tenth the velocity of light) are called fast neutrons See CHAIN REACTION NUCLEAR, FISSION, NUCLEAR, NEUTRON, NEUTRON DELAYED

Neutrons slow down through collisions with nuclei of the surrounding material This slowing down process is made more effective by the introduction of lightweight materials called moderators, such as

heavy water (deuterium oxide), ordinary (light) water, graphite, beryllium, beryllium oxide, hydrides, and organic materials (hydrocarbons) Neutrons that have slowed down to an energy state in equilibrium with the surrounding material are called thermal neutrons The probability that a neutron will cause the fuel material to fission is greatly enhanced at thermal energies, and thus most reactors built utilize a moderator for the conversion of fast neutrons to thermal neutrons

With suitable concentrations of the fuel material, neutron chain reactions can be sustained at higher neutron energy levels The energy range between fast and thermal is designated as intermediate Fast reactors do not have moderators and are relatively small

Reactors have been built in all three categories The first fast reactor was the Los Alamos assembly called Clementine, which operated from 1946 to 1953 The fuel core consisted of nickel coated rods of pure plutonium metal, contained in a 6-in diameter mild (low carbon) steel pot The Dounreay (Scotland) fast reactor has a fuel core in the form of a hexagonal prism, 21 by 21 in, consisting of hollow rods of uranium with 40% uranium 235 It is designed to operate at a heat output of 60 megawatts (Mw), yielding 15 Mw of electrical power An example of an intermediate reactor is the first propulsion reactor for the submarine USS *Seawolf* The fuel core consisted of enriched uranium with beryllium as a moderator, the original coolant was sodium Examples of thermal reactors are given later

Fuel composition Only three isotopes—uranium 235, uranium 233, and plutonium 239—are feasible as fission fuels However, a very wide selection of materials incorporating these isotopes is available *See* PLUTONIUM, URANIUM

Naturally occurring uranium contains only 0 7% of the fissionable isotope uranium 235 the balance being essentially uranium 238 Uranium with higher concentrations of uranium 235 is called enriched uranium Developmental programs for attaining long lived solid fuel elements include studies with metallic uranium, uranium alloys and dispersions (discrete particles of fuel in an inert matrix), and uranium oxide and other refractory uranium compounds Plutonium as a fuel is at a relatively early stage of development

Uranium 233 like plutonium does not occur naturally, but is produced by neutron absorption in thorium 232, a process similar to that by which plutonium is produced from uranium 238 Interest in U^{233} arises from its favorable nuclear properties and abundance of thorium However, studies of this fuel cycle are also at a relatively early stage

Fuel distribution. Fuel moderator assemblies may be homogeneous or heterogeneous Homogeneous assemblies include the aqueous solution type water boilers, molten salt solution dispersions, slurries, and suspensions. Liquid metal fuel reactors may be of the homogeneous or heterogeneous type In the heterogeneous assemblies, the fuel and moderator form separate solid or liquid phases, such as solid fuel elements spaced either in a graphite matrix or in a water phase.

Coolant. The major portion of the energy released by the fissioning of the fuel is in the form of kinetic energy of the fission fragments, which in turn, is converted into heat through the slowing down and stopping of the fragments Heating also arises through the release and absorption of the radiations from the fission process and from the radioactive materials formed

The operating temperature of a reactor is controlled through the removal of heat by a coolant Typical coolants are air, carbon dioxide, light and heavy water, organic substances, and liquid metals In homogeneous reactors, the fuel solution is usually circulated through a heat exchanger external to the reactor

Power rating. The power rating of a reactor is usually given in kilowatts (kw) or megawatts (Mw) of heat The net output of electricity of a nuclear power plant is usually 25–30% of the heat output. The generation of electric energy requires the use of heat to produce steam or to heat gases to drive turbogenerators Thus, the nuclear power plant is quite similar to the conventional coal fired plant except that the nuclear reactor replaces the conventional boiler

Thermoelectric power. In early 1959 the AEC Los Alamos laboratory announced the first successful production of electricity directly from a reactor core without the use of a heat transfer medium or conventional generating equipment. The experimental unit operated by means of a thermoelectric process The thermoelectric medium was cesium vapor, and the heat source was enriched uranium. However this is not expected to be an economic source of electricity in the near future. For additional information *see* THERMOELECTRICITY

CLASSIFICATION BY APPLICATION

Reactor applications include production of plutonium, mobile stationary, and packaged power plants research testing teaching demonstration, and experimental facilities space and process heat and dual purpose designs The potential use of reactor radiation for sterilization of food and other products and for chemical processes is also recognized

Production reactors. Reactor installations at Hanford Washington, Savannah River, South Carolina and Sellafield England are designed to produce plutonium 239 from uranium 238 Natural uranium is used for the fuel material, and the moderator for the reactors at Hanford and at Sellafield is graphite Heavy water is the moderator for the Savannah River reactors The Sellafield reactors (shut down in 1957 following an accident) utilized air as a coolant whereas water is used as a coolant for the United States production reactors. The thermal heterogeneous, natural uranium, graphite-moderated reactors are representative of the very largest sized reactors.

ie term *converter* is used for a reactor that erts a fertile material (for example, uranium to a fissionable material (for example, pluto-i) A breeder reactor, strictly speaking uces the same fissionable material that it umes (for example it consumes plutonium and at the same time breeds plutonium) The cycle, of course, could be based on fissionable nium 233 and fertile thorium 232 rather than nium 235 and plutonium In popular usage, how ;, any reactor that has a conversion ratio of r 100% (that is, produces more fuel than it sumes) is sometimes called a breeder, even if fuels that are consumed and produced are dif ent. This follows from the present lack of plu ium and uranium 233 for reactor fuel

Power reactors Propulsion or mobile nuclear wer reactors are being used for submarines and surface vessels and are being developed aircraft Applications to space vehicles are at early stage of conception and testing The prototype of the first reactor used for propulsion op ated in 1953 and the first reactor powered sub-arine the USS *Nautilus*, was placed in operation 1955 A water coolant is used and is maintained 2000 psi to suppress boiling Pressurized water eactors are in use and are being developed for ibmarines cruisers aircraft carriers merchant ups, and (in the Soviet Union) icebreakers easibility and developmental programs are in rogress for gas cooled, organic moderated and ... ctors for ... le and a ... See Nu CLEAR AIRCRAFT PROPULSION NUCLEAR ROCKET

Central station power reactors have been developed and operated using pressurized water [(PWR) Shippingport Pennsylvania 230 Mw (heat) 60 Mw (electrical) Atomic Power Station (APS) near Moscow U S S R 30 Mw (heat) 5 Mw (electrical)] and gas cooling (Calder Hall reactors, England) Several larger power reactors are nearing completion in the United States, the United Kingdom Canada France, Germany Italy and the Soviet Union

American reactors in use include pressurized water reactors, boiling water reactors fast breeders organic cooled and moderated reactors liquid sodium-cooled reactors and homogeneous recirculating type reactors

The French G-2 and G-3 reactors are thermal, heterogeneous natural uranium graphite-moderated carbon-dioxide-cooled reactors operating at about 200 Mw (heat) and 40 Mw (electrical)

The British power plants at Calder Hall and those scheduled for the Berkeley Bradwell Hunterston and Hinkley Point power stations also are thermal, heterogeneous, natural uranium graphite-moderated, carbon-dioxide-cooled reactors For ex

... ite matrix. The fuel elements are 1¼ in in diameter and 3 ft long and are canned in a magnesium alloy sheath Each fuel channel contains eight fuel elements The fuel elements are provided with longitudinal fins and are cooled with carbon dioxide pressurized to 185 psi The exit gas temperature is 700°F The carbon dioxide coolant transfers its heat to generate steam for the steam turbogenerators, producing about 200 Mw of electrical power

Reactor capacities from 100 to 250 Mw have been planned for power plants in the Soviet Union Types include water cooled reactors with nuclear super heat (that is, with the steam superheated in the reactor by nuclear energy) and use of carbon dioxide at 70 atm pressure and an outlet temperature of 500°C

Packaged nuclear power plants offer unique advantages for remote and isolated installations with reference to transportation and fuel supply Units can be made to be air transportable with long service life per fuel loading

Research reactors. The research and development aspects of a nuclear reactor may be considered from two points of view One is that the reactor provides experimental irradiation facilities and the other is that the reactor itself may represent a test of a given design.

Research with reactors covers such activities as measurements of the probabilities of nuclear reactions, shielding measurements, studies of the behavior of materials under neutron and γ irradiation and other studies in nuclear physics, solid state physics, and in the life sciences The irradiation facilities are used extensively for production of isotopes High neutron flux reactors designed specifically for experimental exposures of materials, are called materials testing reactors Reactors built to test design features are called experiments or experimental reactor facilities Several different types of low cost reactors called teaching demonstration reactors have been promoted to accentuate the teaching aspects

The four major varieties of research reactors are (1) uranium fueled graphite-moderated air cooled reactors, (2) uranium fueled heavy water moderated reactors, (3) enriched fuel aqueous-

...

are heterogeneous Both natural and enriched uranium are used in the first two types

The Bulk Shielding Facility [(BSR) Oak Ridge, Tennessee], was the first reactor with the core submerged in an open pool of water—hence the term swimming pool reactor The BSR fuel is in the form of curved plates each 3 in wide 24 in long, and 0 06 in thick, with 18 plates assembled to form a box of parallel plates The plates consist of enriched uranium (greater than 90%) encased in aluminum sandwiches and clad with aluminum alloy The water spacing between plates is 0 117 in The water is the moderator, coolant and shield With forced circulation of water, reactor levels of

1000 kw of heat are possible Some reactor designs involve the use of a tank instead of a pool Features of other pool and tank type reactors include variation in the fuel element design, fixed and movable cores, and the use of a lightly pressurized (for tank type), forced convection water cooling system

The Material Testing Reactor (MTR) is a high flux irradiation facility designed for studying the behavior of materials for use in power reactors The maximum neutron fluxes available at 40 Mw (heat) are 5.5×10^{14} thermal neutrons/(cm^2) (sec), and 3×10^{14} fast neutrons/(cm^2) (sec) Nearly 100 experimental and instrument holes or exposure ports are provided The fuel assembly is a 3 by 3 by 24-in box containing 19 curved, vertical fuel plates The plates consist of 18% uranium (about 90% enrichment) alloyed with aluminum, in 0 5 mm thick sheets sandwiched between 0 5 mm aluminum sheets Spacing between the plates is 3 0 mm The reactor core consists of fuel elements in 3 by 9 to 5 by 9 arrays Light water is used as a moderator and coolant, and the fuel core is contained in a 54-in diameter aluminum tank Numerous test reactors have been built to accommodate the specialized materials development programs necessary for the expansion of the nuclear reactor industry

Experimental reactors. A variety of reactors have been built merely to test the feasibility of given reactor designs Homogeneous Reactor Experiment No 1 [(HRE 1) Oak Ridge, Tennessee] operated from 1952 to 1954 and was built to demonstrate the potential value and self stabilizing characteristics of a recirculating fuel type reactor Enriched uranyl sulfate (greater than 90%), in *concentrations of 35g per kg of* H_2O, *was pumped* from an 18 in diameter stainless steel sphere to an external heat exchanger The solution was pressurized to 1000 psi, and the maximum fuel solution temperature was 482°F HRE 2, or Homogeneous Reactor Test (HRT), is a 5 Mw recirculating fuel type reactor utilizing a dilute solution of 90% enriched uranyl sulfate in heavy water The fuel solution is pressurized at 2000 psi, and the maximum fuel solution temperature is about 300°C A leak in the core container in 1958 has prevented full operation of the reactor as a two region reactor (One region is the central fuel core and the other is the surrounding heavy water blanket region to be used for breeding or conversion)

First tests on conversion or breeding were carried out in the Experimental Breeder Reactor No 1 (EBR 1), and conversion rates close to unity were attained EBR 1 is a fast reactor with enriched uranium fuel rods and a sodium potassium alloy coolant It was also the first reactor to produce electrical power (200 kw produced with 400 psi steam generation) EBR 2, scheduled to operate in 1960, is designed for 62 5 Mw (heat) and 20 Mw (electrical) The fuel is 49% enriched uranium "fissium" alloy (95% U, 2 5% Mo, 1 5% Ru, 0 3% Rh, 0 5% Pd, 0 2% Zr)

Several types of reactors have been designed and operated under severe power excursions to study reactor stability Four boiling water reactor experiments (Borax 1, 2, 3, and 4) have been performed to study the behavior of boiling water reactors at atmospheric and at elevated pressures, and with different kinds of fuel elements, including nonmetallic fuels Power excursion experiments have been made with the homogeneous aqueous-solution type reactors For example, Kinetic Experiment Water Boiler [(KEWB) Canoga Park, California] has successfully handled a power excursion from 0–530 Mw in less than 1 sec.

The use of boiling water as a coolant for power-producing reactors has been established by the Experimental Boiling Water Reactor [(EBWR) Argonne National Laboratory, Lemont, Illinois] and the Vallecitos Boiling Water Reactor [(VBWR) Vallecitos, California] These test reactors operate at 600 psig and 1000 psig, respectively

The use of sodium as a high temperature coolant for power reactors has been demonstrated by the Sodium Graphite Reactor Experiment (SRE) The SRE is a 21 Mw (heat) and 6 Mw (electrical) thermal, heterogeneous, enriched reactor Graphite is the moderator, and the fuel is in the form of 2 778% enriched uranium slugs, 6 in long by 0 750 in , jacketed with stainless steel tubes.

Organics with sufficient resistance to degradation by irradiation can be used as a coolant or coolant moderator for reactors In the Organic Moderated Reactor Experiment (OMRE), the fuel consists of plates of enriched UO_2 The organic is a polyphenyl compound consisting of 16 8% diphenyl 45 9% *o* terphenyl, 31 5% *m* terphenyl and 5 8% *p*-terphenyl The coolant leaves the reactor at about *710°F*

Process heat. Industry consumes more energy as heat than as electricity The most immediate market available to reactor heat is in supplying low-temperature process steam (about 400°F) to industries such as textiles, paper petroleum, and

supply high temperature heat (over 1500°F) used in specialized chemical and metallurgical applications

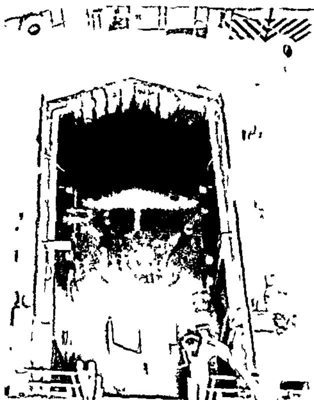
erating electricity or as process heat. Plutonium breeding will be particularly attractive as long as there is a demand for nuclear weapons It is questionable if this leads to optimum efficiency since each purpose must decrease the reactor's capability in performing the other Another possibility for such operation is use of the reactor s radiation for food or chemical processing [H S.I.]

Bibliography C F Bonilla (ed), *Nuclear Engineering*, 1957, J R Dietrich and W H Zinn, *Solid Fuel Reactors*, 1958, H Etherington, *Nuclear Engineering Handbook*, 1958, S Glasstone *Princ*

Low Intensity Testing Reactor at Oak Ridge shown in operation Photograph was taken by using the radiation emanating from the reactor core as the sole source of illumination (*Oak Ridge National Laboratory*)

Bulk Shielding Reactor (BSR) at Oak Ridge It is a so called swimming pool reactor The blue glow given off by the operating reactor is known as Cerenkov radiation (*Oak Ridge National Laboratory*)

Oak Ridge Research Reactor (ORR) a water-cooled and moderated high flux reactor It provides test facilities for research on reactor fuels materials and components under actual conditions of reactor operation (*Oak Ridge National Laboratory*)



1000 kw of heat are possible Some reactor designs involve the use of a tank instead of a pool Features of other pool and tank type reactors include variation in the fuel element design fixed and movable cores and the use of a lightly pressurized (for tank type), forced convection water cooling system

The Material Testing Reactor (MTR) is a high flux irradiation facility designed for studying the behavior of materials for use in power reactors The maximum neutron fluxes available at 40 Mw (heat) are 5.5×10^{14} thermal neutrons/(cm²)(sec), and 3×10^{14} fast neutrons/(cm²)(sec) Nearly 100 experimental and instrument holes or exposure ports are provided The fuel assembly is a 3- by 3- by 24-in box containing 19 curved, vertical fuel plates The plates consist of 18% uranium (about 90% enrichment) alloyed with aluminum in 0.5 mm thick sheets sandwiched between 0.5 mm aluminum sheets Spacing between the plates is 3.0 mm The reactor core consists of fuel elements in 3-by 9 to 5 by 9 arrays Light water is used as a moderator and coolant and the fuel core is contained in a 54-in diameter aluminum tank Numerous test reactors have been built to accommodate the specialized materials development programs necessary for the expansion of the nuclear reactor industry

Experimental reactors. A variety of reactors have been built merely to test the feasibility of given reactor designs Homogeneous Reactor Experiment No 1 [(HRE 1) Oak Ridge Tennessee] operated from 1952 to 1954 and was built to demonstrate the potential value and self stabilizing characteristics of a recirculating fuel type reactor Enriched uranyl sulfate (greater than 90%) in concentrations of 35g per kg of H_2O was pumped from an 18-in diameter stainless steel sphere to an external heat exchanger The solution was pressurized to 1000 psi and the maximum fuel solution temperature was 482°F HRE 2 or Homogeneous Reactor Test (HRT) is a 5 Mw recirculating fuel type reactor utilizing a dilute solution of 90% enriched uranyl sulfate in heavy water The fuel solution is pressurized at 2000 psi and the maximum fuel solution temperature is about 300°C A leak in the core container in 1958 has prevented full operation of the reactor as a two-region reactor (One region is the central fuel core and the other is the surrounding heavy water blanket region to be used for breeding or conversion)

First tests on conversion or breeding were carried out in the Experimental Breeder Reactor No 1 (EBR 1) and conversion rates close to unity were attained EBR 1 is a fast reactor with enriched uranium fuel rods and a sodium potassium alloy coolant. It was also the first reactor to produce electrical power (200 kw produced with 400 psi steam generation) EBR 2, scheduled to operate in 1960 is designed for 62.5 Mw (heat) and 20 Mw (electrical) The fuel is 49% enriched uranium "fissium" alloy (95% U 2.5% Mo 1.5% Ru 0.3% Rh, 0.5% Pd, 0.2% Zr)

Several types of reactors have been designed and operated under severe power excursions to study reactor stability Four boiling water reactor experiments (Borax 1, 2 3 and 4) have been performed to study the behavior of boiling water reactors at atmospheric and at elevated pressures, and with different kinds of fuel elements, including nonmetallic fuels Power excursion experiments have been made with the homogeneous aqueous solution type reactors For example, Kinetic Experiment Water Boiler [(KEWB) Canoga Park, California] has successfully handled a power excursion from 0–530 Mw in less than 1 sec

The use of boiling water as a coolant for power-producing reactors has been established by the Experimental Boiling Water Reactor [(EBWR) Argonne National Laboratory, Lemont, Illinois] and the Vallecitos Boiling Water Reactor [(VBWR) Vallecitos, California] These test reactors operate at 600 psig and 1000 psig respectively

The use of sodium as a high temperature coolant for power reactors has been demonstrated by the Sodium Graphite Reactor Experiment (SRE) The SRE is a 21 Mw (heat) and 6 Mw (electrical) thermal heterogeneous enriched reactor Graphite is the moderator and the fuel is in the form of 2.778% enriched uranium slugs 6 in long by 0.750 in , jacketed with stainless steel tubes

Organics with sufficient resistance to degradation by irradiation can be used as a coolant or coolant moderator for reactors In the Organic Moderated Reactor Experiment (OMRE), the fuel consists of plates of enriched UO_2 The organic is a polyphenyl compound consisting of 16.8% diphenyl 45.9% *o*-terphenyl 31.5% *m* terphenyl and 5.8% *p*-terphenyl The coolant leaves the reactor at about 710°F

Process heat Industry consumes more energy as heat than as electricity The most immediate market available to reactor heat is in supplying low temperature process steam (about 400°F) to industries such as textiles paper petroleum, and

supply high temperature heat (over 1500°F) used in specialized chemical and metallurgical applications.

Dual purpose reactors It is technically feasible to use a reactor for breeding plutonium or uranium-233 while at the same time using the heat for generating electricity or as process heat Plutonium breeding will be particularly attractive as long as there is a demand for nuclear weapons It is questionable if this leads to optimum efficiency since each purpose must decrease the reactor's capability in performing the other Another possibility for such operation is use of the reactor's radiation for food or chemical processing [H.S.I.]

Bibliography C. F Bonilla (ed) *Nuclear Engineering* 1957 J R Dietrich and W H Zinn, *Solid Fuel Reactors*, 1958 H Etherington, *Nuclear Engineering Handbook* Glasstone

Low Intensity Testing Reactor at Oak Ridge shown in operation. Photograph was taken by using the radiation emanating from the reactor core as the sole source of illumination. (*Oak Ridge National Laboratory*)

Bulk Shielding Reactor (BSR) at Oak Ridge. It is a so-called swimming pool reactor. The blue glow given off by the operating reactor is known as Cerenkov radiation. (*Oak Ridge National Laboratory*)

Oak Ridge Research Reactor (ORR), a water-cooled and moderated high flux reactor. It provides test facilities for research on reactor fuels, materials and components under actual conditions of reactor operation. (*Oak Ridge National Laboratory*)



ples of Nuclear Engineering, 1955; S Glasstone, *Sourcebook on Atomic Energy, 2d ed*, 1958; A. W. Kramer, *Boiling Water Reactors*, 1958; R L. Murray, *Introduction to Nuclear Engineering*, 1954; C Starr and R W Dickinson, *Sodium Graphite Reactors*, 1958, R Stephenson, *Introduction to Nuclear Engineering*, 2d ed, 1958, U S. Atomic Energy Commission, Westinghouse Electric Corp, and Duquesne Light Co, *The Shippingport Pressurized Water Reactor*, 1958

Reactor, ship propulsion

Nuclear reactors for shipboard propulsion can, in theory, be of any type used for the production of useful power For basic information applicable to all types of fission reactors, see REACTOR, NUCLEAR, REACTOR NUCLEAR (CLASSIFICATION), REACTOR PHYSICS

From a practical standpoint, the differences between shipboard applications and other applications are substantial Shipboard nuclear power plants must contain (1) a nuclear reactor which constitutes a heat source, (2) some form of energy transfer system (usually called the coolant system) to convey the heat energy released in the nuclear reactor to the power plant system itself, and (3) a means for converting this energy into useful mechanical work that is, for turning a propeller shaft and generating electricity for auxiliary purposes In all of the shipboard nuclear power plants which have been built or which are known to be under construction, the energy conversion step is based on the steam turbine cycle, and that portion of the plant is more or less conventional [illegible]

[illegible] ferences between shipboard reactors and similar installations ashore, involving (1) problems of weight and space limitations, (2) problems of plant reliability and onboard maintenance, (3) problems in plant safety, and (4) problems inherent in location on a moving platform.

Weight and space problems The actual size of the shipboard reactor itself does not present a problem, but the size and configuration of the shielded volume around the entire cooling system are of considerable significance. Figure 1 shows a simplified schematic diagram of a typical marine nuclear propulsion plant. Both the primary shield around the pressure vessel containing the nuclear reactor itself and the secondary shield located around all radioactive components of the cooling system are shown For a discussion of the theory and application of reactor shielding, see RADIATION SHIELDING.

Weight considerations usually result in a compact layout to reduce the size and hence the weight of the secondary shield However, the conflicting requirements for access and maintenance result in a balance being required such that the permeability (ratio of free volume of space to total volume

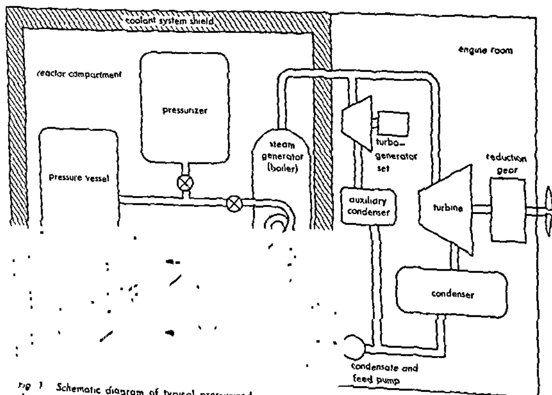


Fig 1 Schematic diagram of typical pressurized nuclear [illegible] lative sizes not indicated) [illegible] ne Engrs]

of space) of the portion of a nuclear power plant within the secondary shield usually runs from 65 to 70% as opposed to approximately 80% permeability for normal shipboard machinery spaces exclusive of uptakes, air vents, or tonnage openings *See* SHIP DESIGN

It will be difficult to reduce the weight of present day shields for pressurized water plants because the principal γ ray attenuation results from the use of heavy materials

Reliability and maintenance problems Nuclear power confers on any vessel to which it is applied an exceptionally long time interval between refuelings Therefore, the importance of plant reliability and the ability to perform reasonable maintenance functions on board ship assume increased importance as compared to conventional shipboard power plants In particular, planning for access to plant components and planning for necessary

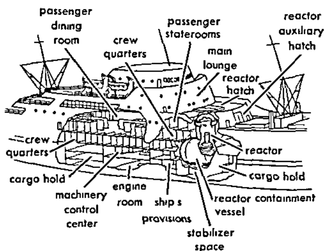


Fig 2 NS *Savannah* cutaway, showing relationship of reactor to ship s spaces (*From Nucleonics, vol 16, no 9, 1958*)

maintenance functions to be performed on these plant components is essential Many of these components will transfer radioactive fluids and therefore special provisions must be made either in the form of installed standby equipment or by the provision of cleaning and servicing equipment

Plant safety problems A nuclear reactor on board ship is considerably more subject to external hazards, with resulting possibility of nuclear accident, than is a similar plant located ashore Special precautions in the form of collision damage protection, secondary containment, and special safety devices are considered necessary Figures 2 and 3 show how the vital parts are protected from collision damage to the greatest extent possible by careful location The principal hazard from a nuclear accident results from the spreading of radioactivity Therefore, any reactor system which has the fission product activity held in a reasonably permanent form inside solid fuel elements and which has a coolant of low long term activity possesses an advantage from a safety standpoint

Another important problem affecting safety is the characteristic of a nuclear reactor that, even

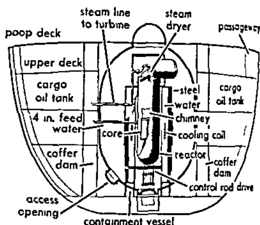


Fig 3 Typical arrangement of boiling water reactor machinery in containment vessel (*Soc Naval Architects Marine Engrs*)

when shut down, it continues to evolve heat at a low rate from residual radioactivity Adequate means must be provided for disposing of this radiation during periods of inoperation and even in cases where the vessel is sunk

Seakeeping problems A ship is subject to substantial motions in waves, thus any nuclear power reactor and its associated equipment must be designed in accordance with all the accepted basic principles of marine engineering If large free surfaces are present, special precautions must be taken to protect them If fluids that are rare and difficult to obtain are used in the reactor system (either as coolants or as auxiliary fluids), provision must be made either for their generation on board or for adequate storage of a reserve supply If high melting point materials are involved preheating systems with adequate energy sources must be provided Exceptionally reliable auxiliary power sources are required because of the safety and control problems involved

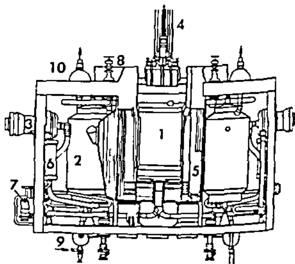


Fig 4 General layout of one reactor unit in *Lenin*. 1 Reactor, 2 steam generator; 3 main circulating pump 4 control rod mechanism; 5 filter; 6 cooler 7 secondary circuit pump; 8 primary steam valve 9 feed water inlet 10 steam outlet (*From Lenin, The Russian icebreaker, Nuclear Eng , 3(31) 432–433 1958*)

luclear powered submarines and surface ships of the U S Navy

Name	Type	Pressurized water reactor designer	Shipbuilder	Status
Submarines				
USS *Nautilus*	Attack	Westinghouse	Electric Boat	Completed 1/22/55
USS *Seawolf*	Attack	General Electric*	Electric Boat	Completed 1/21/57
USS *Skate*	Attack	Westinghouse	Electric Boat	Completed 12/20/57
USS *Swordfish*	Attack	Westinghouse	Portsmouth	Completed 9/15/58
USS *Sargo*	Attack	Westinghouse	Mare Island	Completed 10/1/58
USS *Seadragon*	Attack	Westinghouse	Portsmouth	Completed 12/5/59
USS *Skipjack*	Attack	Westinghouse	Electric Boat	Completed 4/15/59
USS *Triton* (2 reactors)	Radar picket	General Electric	Electric Boat	Completed 11/10/59
USS *Halibut*	Regulus missile	Westinghouse	Mare Island	Completed 1/4/60
Scamp	Attack	Westinghouse	Mare Island	Under construction
Scorpion	Attack	Westinghouse	Electric Boat	Launched 12/19/59
Sculpin	Attack	Westinghouse	Ingalls	Under construction
Shark	Attack	Westinghouse	Newport News	Under construction
Snook	Attack	Westinghouse	Ingalls	Under construction
Thresher	Attack	Westinghouse	Portsmouth	Under construction
Permit	Attack	Westinghouse	Mare Island	Under construction
Plunger	Attack	Westinghouse	Mare Island	Under construction
	Attack	Westinghouse	Ingalls	Under construction
	Hunter killer	Combustion Engineering	Electric Boat	Under construction
rge Washington	Polaris missile	Westinghouse	Electric Boat	Completed 12/30/59
Henry	Polaris missile	Westinghouse	Electric Boat	Launched 9/18/59
Roosevelt	Polaris missile	Westinghouse	Mare Island	Launched 10/22/59
Lee	Polaris missile	Westinghouse	Newport News	Launched 12/18/59
Lincoln	Polaris missile	Westinghouse	Portsmouth	Under construction
	Attack	Westinghouse	N Y Shipbuilding	Under construction
	Attack	Westinghouse	N Y Shipbuilding	Under construction
	Attack	Westinghouse	Portsmouth	Under construction
	Attack	Westinghouse	Portsmouth	Under construction
	Attack	Westinghouse	Ingalls	Under construction
Allen	Polaris missile	Westinghouse	Electric Boat	Under construction
Houston	Polaris missile	Westinghouse	Newport News	Under construction
s A Edison	Polaris missile	Westinghouse	Electric Boat	Authorized
Marshall	Polaris missile	Westinghouse	Newport News	Authorized
assigned	4 attack	4 not assigned	4 not assigned	4 authorized
e ships				
d missile cruiser g Beach (2 reactors)		Westinghouse	Bethlehem	Launched 7/14/59
aft carrier *Enterprise* reactors)		Westinghouse	Newport News	Under construction
ed missile destroyer inbridge (2 reactors)		General Electric	Bethlehem	Under construction

Originally built with General Electric–designed sodium-cooled reactor replaced by Westinghouse pressurized er reactor

Reactor types Only two types of reactor, the ssurized water reactor and the sodium cooled ctor have actually been applied to operating sels The pressurized water plant has been the orite because it can be more easily shielded and intained in working order Other nuclear power oolants all have ion and (1955– 959) Even the sodium cooled reactor originally nstalled on the submarine USS *Seawolf* has been eplaced by a pressurized water reactor

All the principal maritime nations are studying the application of nuclear power to naval and commercial ships. The United States Great Britain and the Soviet Union were the first nations actually to construct nuclear vessels

The United States Navy has in operation and under construction a large number of nuclear powered submarines and surface vessels, as shown in the accompanying table (for additional information on some of these vessels see SHIP, NAVAL, SUBMARINE) The first commercial marine installations of nuclear power are on the Soviet ice breaker *Lenin* (Fig 4) and on the United States merchant ship N S *Savannah*, both of which use the pressurized water type of reactor The *Lenin* was launched on December 5 1957, and the *Savannah* was launched on July 21, 1959 [L H R]

Bibliography The atomic navy, *Nucleonics*, August, 1957, and September 1959, C Hinton and

R V Moore The nuclear propulsion of ships *Trans Inst of Marine Engrs*, 1957, J W Landis, The power plant for the first nuclear merchant ship NS *Savannah*, *J Am Naval Engrs*, 70(4) 629–641, 1958, H G Rickover, J M Dunford, T Rockwell, W C Barnes, and M Shaw Some problems in the application of nuclear propulsion to naval vessels *Soc Naval Architects Marine Engrs Trans* 65, 1958, T Rockwell (ed), *Reactor Shielding Design Manual*, 1956 L H Roddis, Jr and J W Simpson, The nuclear propulsion plant of the USS *Nautilus*, SSN 571, *Soc Naval Architects Marine Engrs Trans*, 62 491–521 1954, H E Vann, M L Weiss, and B Wolfe, Shielding aspects of nuclear power plants for marine propulsion *Soc Naval Architects Marine Engrs Trans*, 66 1958

Reactor physics

The science of the interaction of the elementary particles and radiations characteristic of nuclear reactors with matter in bulk These particles and radiations include neutrons beta (β) rays and gamma (γ) rays of energies between zero and about 10^7 electron volts (ev) *See* BETA RAYS GAMMA RAYS, NEUTRON REACTOR NUCLEAR

The study of interaction of β and γ radiations with matter is, within the field of reactor physics, undertaken primarily to understand the absorption and penetration of energy through reactor shields For a discussion of problems of reactor shielding *see* RADIATION SHIELDING

With this exception reactor physics is the study of those processes pertinent to the chain reaction involving neutron induced nuclear fission with consequent neutron generation Reactor physics is differentiated from nuclear physics which is concerned primarily with nuclear structure Reactor physics makes direct use of the phenomenology of nuclear reactions Neutron physics is concerned primarily with interactions between neutrons and individual nuclei or with the use of neutron beams as analytical devices whereas reactor physics considers neutrons primarily as fission producing agents In the hierarchy of professional classification neutron physics and reactor physics are both ranked as subfields of the more generalized area of nuclear physics *See* FISSION NUCLEAR NUCLEAR PHYSICS

Reactor physics borrows most of its basic concepts from other fields From nuclear physics comes the concept of the nuclear cross section for neutron interaction defined as the effective target area of a nucleus for interaction with a neutron beam The total interaction is the sum of interactions by a number of potential processes and the probability of each of them multiplied by the total cross section is designated as a partial cross section Thus a *given nucleus is characterized by* cross sections for capture fission elastic and inelastic scattering and also for such reactions as $(n\ p)$, $(n\ \alpha)$, and $(n\ 2n)$ An outgrowth of this is the definition of macroscopic cross section which is the product of cross section (termed microscopic for specificity) with atomic density of the nuclear species involved The symbols $N\sigma_i^j$ or Σ_i^j are used for macroscopic cross section the subscript referring to the nuclear reaction involved and the superscript to the isotope. The dimensions of Σ are cm^{-1}, and the usual magnitude of Σ_t (total cross section) is of the order of magnitude of unity

Cross sections vary with energy according to the laws of nuclear structure In reactor physics, this variation is accepted as input data to be translated into a description of neutron behavior Common aspects of cross section dependence, such as variation of absorption cross section inversely as the square root of neutron energy or the approximate regularity of resonance structure form the basis of most simplified descriptions of reactor processes in terms of mathematical or logical models

The concept of neutron flux is related to that of macroscopic cross section This may be defined as the product of neutron density and neutron speed, or as the rate at which neutrons will traverse the outer surface of a sphere imbedded in the medium per unit of spherical surface The units of flux are neutrons/(cm^2) (sec) The product of flux and macroscopic cross section yields the reaction rate per unit volume and time The use of these variables is conventional in reactor physics

The chain reaction a concept derived from chemical kinetics is the basis of a physical description of the reactor process The source of energy in a nuclear reactor is the fission of certain isotopes of heavy elements (thorium uranium, and plutonium in particular) when they absorb neutrons Fission splits the elements into two highly radioactive fragments which carry away most of the energy liberated by the fission process (about 160 Mev out of about 200 Mev per fission) the bulk of which is rapidly transformed into heat. Neutrons of high energy are also liberated so that a chain of events alternating between neutron production in fission and neutron absorption causing fission may be initiated Because more than one neutron is liberated per fission this chain reaction may rapidly branch out to produce an increasing reaction rate or divergent reaction or if the arrangement of materials is such that only a small fraction of the neutrons will ultimately produce fission the chain will be broken in a convergent reaction

Criticality The critical condition is what occurs when the arrangement of materials in a reactor allows on the average exactly one neutron of those liberated in one nuclear fission to cause one additional nuclear fission If a reactor is critical it will have fissions occurring in it at a steady rate This *desirable condition is achieved by balancing the* probability of occurrence of three competing events fission neutron capture in a material which does not undergo fission and leakage of neutrons from the system If ν is the average number of neutrons liberated per fission then criticality r

...nder wh ch the probability of a neu

k_e the ratio of fissions in success ... chain, as a product of probabilities of successive ...

those used to characterize nuclear even ... those used to characterize leakage In the former group are fast effect ϵ resonance escape probability p thermal utilization factor f and neutrons emitted per fuel absorption η In the latter group are neutron age or slowing down area τ or L_s^2 migration area M^2 thermal diffusion area L^2 diffusion coefficient D and buckling B^2

Fast effect The fast effect occurs in thermal reactors containing significant quantities of U^{238} or Th^{232} These reactors comprise the bulk of plutonium production and civilian power reactors The isotopes mentioned can undergo fission only when struck by very energetic neutrons only a little more than one-half of the neutrons born in fission can cause fission in them Moreover these fast neutrons (energy greater than about 1 4 Mev) are subject to energy degradation by the competing reaction of inelastic scattering and by collision with moderator In consequence only a small number of fast fissions will occur

The fast effect is characterized by a ratio R of ...ist to nonfast fissions From this ratio a quantity ... the fast effect is derived which determines the ...eutrons available to the chain reaction after the ...onvergent chain of fast fissions has been com...pleted per neutron born of nonfast fission

The common magnitude of ϵ varies between ...02 and 1 05 R varies between 0 and 0 15 commonly When $R < 0\ 01$ the fast effect is usually ...omitted from consideration

Resonance escape probability The resonance escape probability p is a significant parameter for the same group of reactors that have a significant fast effect and is defined as the probability that a neutron in the course of being moderated will escape capture by U^{238} (or Th^{232}) at any of several energies at which the capture cross section is unusually high (resonance energies) The existence of this resonance absorption prevents most reactors homogeneously fueled with natural uranium from going critical because resonance absorption coupled with other losses causes neutron depletion below the requirement for criticality Therefore low enrichment reactors are heterogeneous the fuel being disposed in lumps The lumping of fuel originally proposed by E Fermi and E P Wigner increases the probability of resonance escape Spatial isolation of resonance absorber from moderator increases the number of neutrons which are slowed down without making any collisions with the absorber Also because absorption is very probable at resonance energ es the neutrons can travel only very short distances into the fuel lump before being absorbed so that the interior of the lump is hardly exposed to resonant neutrons Another effect of lumping is the removal of excess neutron scattering near the absorber so that there is a lesser probability of absorption following multiple collision

With slight enrichment homogeneous assemblies can be made critical The homogeneous problem also provides the formalism by which p can be calculated Use is made of the resonance integral RI defined as the absorption probability per absorbing nucleon per neutron slowed down from infinite source energy in a moderator of unit slowing-down power In a highly dilute system

$$RI = \int_{E_0}^{\infty} \sigma_a \, dE/E$$

where σ_a is the neutron absorption cross section of the absorber as a function of energy and E_0 is an energy taken as the lower limit of the resonance region The slowing down power of a moderator nucleus is $\xi\sigma_s$ where ξ is the mean increase in lethargy of the neutron per scattering and σ_s is the moderator scattering cross section Thus the resonance absorption probability of a system would appear to be

$$1 - p = N\ (RI)/N_m\xi\sigma_s$$

where N and N_m are respectively atomic concen...

...plied a better expression is

$$p = \exp\{-N\ (RI)/N_m\xi\sigma_s\}$$

The value of E_0 used in the definition of RI follows one of two conventions it is either taken as 0 4 ev the energy below which thin sheets of cadmium absorb all neutrons or as an energy just below the lowest resonance about 6 ev for U and 20 ev for Th In the latter case RI is spoken of as $1/v$ corrected

Thermal utilization factor Thermal utilization factor f is the fraction of neutrons which once thermalized are absorbed in fuel In a homogeneous array f may be calculated from the atomic densities (that is the number of atoms per cubic centimeter) and thermal absorption cross sections of the various constituents of the reactor The problem becomes more complex in a highly absorbing system and in the presence of nuclei (such as Pu or Cd) whose absorption cross sections do not vary with energy in the usual $1/v$ fashion In this case it becomes necessary to evaluate the neutron spectrum and average the absorption cross sections over this spectrum in order to obtain reaction ratios

For low absorptions the neutron spectrum is given by the Maxwellian expression

$$N(E)\, dE = [2^{\pi}/(\pi kT)^{3/2}]\quad E^{1/2} \exp -(E/kT)\, dE$$

where k is the Boltzmann constant and T the absolute temperature of the medium Deviations from

this shape caused by absorption were first formulated by E P Wigner and J E Wilkins This problem has recently become highly significant with the advent of highly absorbing systems operating with considerable quantities of Pu

In heterogeneous systems, the problem is further complicated by the spatial nonuniformity of the neutron flux It is therefore necessary to calculate reaction rates in various regions of the lattice by multiplying local absorption cross sections and atomic densities by local neutron fluxes, or, in effect the same thing, by introducing flux weights into the cross sections

The calculation of neutron flux may be performed by diffusion theory or by more exact and elaborate methods for solving the neutron transport problem The problem may be further complicated by spatial effects on spectrum

The term disadvantage factor, applied in simple systems originally to describe the ratio of mean moderator to mean fuel flux, has fallen into disfavor because of vague and local definitions The symbol F, called the fuel disadvantage factor, is still in use to describe the ratio of surface to mean volume flux in a fuel lump with isoperimetric flux *See* THERMAL NEUTRONS

Fission neutrons per fuel absorption This constant, η, is a characteristic of the fuel and of the neutron spectrum, but not of the spatial configuration of the system Pure fissionable materials do not always undergo fission when they absorb neutrons, sometimes they lose their energy of excitation by emission of a gamma ray The number of neutrons per fission ν varies only slightly with incident neutron energy (a few per cent per million electron volts), but η can fluctuate considerably even within a fraction of an electron volt Consequently, the specification of η is dependent upon a good evaluation of neutron spectrum

For unirradiated, low enrichment assemblies the custom of defining fuel as all uranium, U^{235} and U^{238}, persists Thus f considers total uranium captures, and η is lowered from the value for pure U^{235} by the fractional absorption rate of U^{235} in uranium This custom leads to excessive complexity as plutonium builds into the fuel, and is therefore declining in use

Infinite multiplication constant The infinite multiplication constant, k_∞, is the ratio of neutrons in successive generations of the chain in the absence of leakage In the formalism just described, the chain is taken from thermal neutron through fission and back to thermal neutron and from the definition of terms, $k_\infty = \eta\epsilon p f$ This formula is known as the four factor equation

For reactors other than weakly absorbing thermal systems, the simplified description breaks down Thus in a fast reactor significant fission and capture occur at all neutron energies, and no moderator is present, in a very strongly absorbing thermal system, an appreciable fraction of neutrons react at energies between the thermal region and the lowest U^{238} resonance For such systems, the definition of the neutron chain is usually made in terms of a total time dependent fission rate expression, and the parametric representation of k_∞ becomes appreciably more complex Generally the spectrum is broken up into energy groups, and k_∞ is defined as the sum of the fission neutron production rate over all groups divided by the sum of absorption rates over all groups

Neutron age The neutron age τ is a reactor parameter defined in various ways, all related to the probability of leakage of a fast neutron from a reactor system The basic definition is that 6τ measures the mean square distance of travel between injection of a neutron at one energy or energy spectrum, and its absorption at some other energy in an infinite system The term age is used because in weakly moderating systems the equation describing neutron slowing down in space and energy has the same form as the time-dependent heat-conduction equation, with τ substituted for time. The British and Canadian usage is L_s^2, for slowing down length (squared), which more accurately describes the physical parameter

The common injection spectrum is a fission spectrum, and the common points of measurement or application are absorption at the indium resonance energy, 1 4 ev, or at some arbitrarily defined thermalization energy The ages of these energies are denoted as τ_{In} or τ_{Th} When a source other than the fission spectrum is considered, other subscripts are used

A M Weinberg has pointed out that the shape of the spatial distribution in an infinite system for which τ is the second moment can be closely correlated with the fast leakage of a bare finite system If the finite system has a source and sink distribution which is a solution of Laplace's equation $\nabla^2\phi + B^2\phi = 0$, then the Fourier transform of the τ distribution for given B will yield a quantity $P_\infty(B^2)$ which is almost exactly the nonleakage probability during slowing down Two particularly significant cases are those for which the τ distribution is a Gaussian or an exponential curve In the former case $P_\infty(B^2) = e^{-\tau B^2}$, in the latter $P_\infty(B^2) = 1/(1 + \tau B^2)$ The Gaussian distribution is experimentally and theoretically verified for moderators as heavy as Be or heavier The exponential distribution is crudely applicable to H_2O moderated systems and D_2O has a definitely mixed distribution

In some cases τ is used in a synthetic way to describe the leakage under some simple approximation to the slowing down distribution Thus, τ_{2G} is a number which is used in a two-energy group neutron model to give correct fast nonleakage probability as $P(B^2) = 1/(1 + \tau_{2G}B^2)$ When the model is not a good approximation, $6\tau_{2G}$ is not the second moment of the τ distribution

For a heavy moderator, the age between two energies is given by the simple approximation

$$\tau_{E_1 \to E_2} = \int_{E_2}^{E_1} D(E)/3(N\xi\sigma_s)\, dE/E$$

where D is the diffusion coefficient N is atomic density σ_s is scattering cross section and ξ is mean lethargy (logarithmic energy) gain/collision

Thermal diffusion area The thermal diffusion area L^2 is one-sixth the mean square distance of travel *between thermalization* and absorption It is thus the analog of τ for thermal neutrons Because thermal migration is usually well represented by a diffusion equation which has an *exponential* absorption distribution thermal nonleakage probability is given by $1/(1 + L^2B^2)$ L^2 is defined by $L^2 = D_{Th}/N\sigma_{a\,Th}$ where D and σ_a are spectrum averaged diffusion coefficient and absorption cross section respectively

Migration area The migration area M^2 is one-sixth the mean square distance of travel from birth to death of a neutron For large small leakage systems $1/(1 + M^2B^2)$ is an excellent approximation *to the total nonleakage* probability When only thermal absorption exists $M^2 = \tau + L^2$

Diffusion coefficient The diffusion coefficient D is essentially a scaling factor applied *to* validate s law a relation between neutron flux and cur (the latter being defined as a vector describing ate of flow of neutron density) Fick s law is essed as

$$\jmath = -D\nabla\phi$$

re $\jmath$ is current, $\nabla\phi$ flux grad ent and D diffusion ficient For generalized systems D is a tensor but egions where $|\nabla\phi|/\phi$ is small D is approximated

$\nabla^2\phi/\phi$
ntity is
name is

ived from the relationship between force and lection in a mechanical system with constrained undaries In a bare reactor the buckling is con nt except within one neutron mean free path $\equiv 1/\Sigma$) of the boundary where spectral ef cts cause perturbation For a slab of width t $= (\pi/t)^2$ *for an infinitely high cylinder of ra*- us a $B^2 = (2\,404/a)^2$ for a sphere of radius a $= (\pi/a)^2$ and for other geometries it is again geometrical constant Because B^2 is a definite genvalue of Laplace s equation it is the appropri te number to be used in the formulations of $P(B^2)$ reviously described

Effective multiplication k_{eff} is the ratio of neu ron production *in successive generations* of the hain reaction It is given by the product of k_∞ and $P(B^2)$ where $P(B^2)$ is itself the product of fast and *thermal* (or equivalent) nonleakage probabili ties

Reactivity Reactivity is a measure of the deviation of a reactor from the critical state at any frozen instant of time The term is qualitative because three sets of units are in *current use to describe* it.

Per cent k and millikay are absolute units describing *the imbalance of the* system from critical ty per fission generation Because $k_{eff} = 1$ describes a critical system one says that it is 1% super or subcritical *respectively if each generation* produces 1 01 or 0 99 times as many neutrons as the preceding one Millikay are units of 0 1% k and are given plus sign for supercriticality and minus sign for subcriticality In both cases k_{eff} is the base

$$\%k = 100\,(k_{eff} - 1)$$
$$\text{Millikay} = 1000\,(k_{eff} - 1)$$

These units are used primarily in design and analysis of *control rods*

Dollars are relative units, describing reactivity in terms of the mean fraction of delayed neutrons per fission Because the delayed neutrons are the primary agents for permitting control of the reac *tion supercriticalities* of less than 1 dollar are con sidered manageable in most cases Thus there is 1 dollar to "spend in maneuvering power level The dollar is subdivided into 100 cents.

Because the delayed neutron fraction is a function of both neutron energy and fissionable material the conversion rate between dollars and %k varies among reactors. Notwithstanding its origins the dollar has found greatest acceptance as *a unit* for analyzing reactor runaways and nuclear explosions where reactivities greater than 1 dollar are considered

Inhours are reactivity units based on rate of change of power level *in low power reactors. If a* low power reactor is given enough reactivity so that its level would steadily increase by a factor of e per hour (which is also known as a 1 hr period) that much reactivity is 1 inhour (from inverse hour) *It is only for very small reactivities* how ever that reactivity in inhours may be obtained from reciprocal periods

Reactivity is measured in inhours primarily by operators of steady state reactors, in which only small reactivities are normally encountered

Reflectors Reflectors are bodies of material placed beyond the chain reacting zone of a reactor *whose function is to return* to the active zone (or core) neutrons which might otherwise leak Reflector worth can be crudely measured in terms of the albedo or probability that a neutron passing from core to reflector will return again to the core

Good reflectors are *materials with high scatter* ing cross sections and low absorption cross sections The first requirement ensures that neutrons will not easily diffuse through the reflector and the second that they will not easily be captured in dif fusing back to the core

Beryllium is the outstanding reflector material in terms of neutronic performance Water graphite D_2O iron lead and U^{238} are also good reflectors The use of Be H_2O C, and D_2O as reflectors permits conversion of neutrons leaking at high energy into thermal neutrons diffusing back to the core Because the reverse flow of neutrons is always accompanied by a *neutron flux gradient* these reflectors show characteristic thermal flux peaks outside the core They are therefore desir

able materials for research reactors in which flux peaks are useful for experimental purposes

The usual measurement of reflector worth is in terms of reflector savings defined as the difference in the reflected dimension between the actual core and one which would be critical without reflector Reflector savings are close to reflector dimensions for thin reflectors, and approach an asymptotic value dependent upon core size and reflector con stitution as thickness increases

Reactor dynamics Reactor dynamics is con cerned with the temporal sequence of events when neutron flux power, or reactivity varies The inclu sive term takes into account sequential events not necessarily concerned with nuclear processes, which may affect these parameters There are basi cally three ways in which a reactor may be affected so as to change reactivity A control element, ab- sorbing rod or piece of fuel may be externally ac tuated to start up shut down, or change reactivity or power level, depletion of fuel and poison buildup of neutron absorbing fission fragments and production of new fissionable material from the fertile isotopes Th^{232}, U^{234} U^{238} and Pu^{240} make reactivity depend upon the irradiation his tory of the system and changes in power level may produce temperature changes in the system lead ing to thermal expansions and mechanical changes of its constituents with consequent change of re activity

Reactor control physics Reactor control phys ics is the study of the effect of control devices on reactivity and power level As such it includes a number of problems in reactor statics because the primary question is to determine the absorption of the control elements in competition with the other neutronic processes It is however a problem in dynamics given the above information to deter mine what motions of the control devices will lead to stable changes in reactor output

Particular problems occurring in the statics of reactor control stem from the particular nature of control devices Many control rods are so heavily absorbing for thermal neutrons that elegant refine ments of neutron transport theory are needed to es timate their absorption Other types of control rods include isotopes with heavy resonance absorp tion and the interaction of such absorbers with U^{238} resonances must be examined The motion of control rods changes the material balance of reac tor regions and with water moderated reactors peaks in the fission rate occur near empty rod channels By virtue of high absorption of their con stituent isotopes some absorbing materials (for example, cadmium and boron) burn out in the re actor and a rod made of these materials loses ab sorbing strength with time As a final example the motion of a control rod may change the shape of the power pattern in the reactor so as to bring sec ondary pseudostatic effects into play *See* REAC TOR NUCLEAR

Reactivity changes Long term reactivity changes may represent a limiting factor in the burning of nuclear fuel without costly reprocess ing and refabrication As the chain reaction pro ceeds, the original fissionable material is depleted and the system would become subcritical if some form of slow addition of reactivity were not avail able This is the function of shim rods in a typical reactor The reactor is originally loaded with enough fuel to be critical with the rods completely inserted As the fuel burns out, the rods are with drawn to compensate

In order to decrease requirements on the shim system, many devices to overcome reactivity loss may be used A burnable poison may be incorpo rated in the system This is an absorbing isotope which will burn out at a rate comparable to or greater than the fuel Burnable poisons are there fore limited to isotopes with very high effective neutron absorption cross sections Combinations of poisons and the use of self shielding of poisons can, in principle make the close compensation of considerable reactivity possible without major con trol rod motions, but the technological problems in their use are formidable

A more popular method for compensating reac tivity losses is the incorporation of fertile isotopes into the fuel This is desirable because the neu trons captured in the fertile material are not wasted but used to manufacture new fissionable

fissionable atoms formed per old fissionable atom burned) and the fission parameters of the materi als a reactor so fueled loses reactivity relatively slowly and in some cases may show a temporary reactivity increase The various isotopes produced by successive neutron capture in uranium and plu tonium must be considered in this problem, the higher isotopes becoming prominent at very long exposures

A final consideration of long term reactivity is the extra parasitic absorption of the fission prod ucts as formed At long exposure this absorption becomes significant because of the relatively high absorption of many of the fission products. At shorter exposures isotopes of very high absorp tion mainly Sm^{149} and Xe^{135} are more promi nent These materials have such high cross sections that they reach a steady state concentration rela tively quickly burning out by neutron capture as rapidly as they are formed in fission

Xe^{135} is particularly interesting because its cross section is abnormally large its fission yield is high it is preceded by an isotope of low cross sec tion and approximately 7 hr half life (I^{135}) and it undergoes β decay to the low absorption Cs with a half life of about 10 hr This combination of properties gives several interesting effects One of these is that at high flux a reservoir of I^{135} is formed which continues to decay to Xe^{135} even after the reactor is shut down Bec

aintained at steady state during operation by a ilance of buildup against burnout the shutdown so removes the chief mode of Xe^{135} destruction ence the Xe^{135} concentration increases immedi tely after shutdown The reservoir of I^{135} is so irge that reactivity is rapidly lost as Xe^{135} builds p and in some cases it may be impossible to re- iart the reactor after a short shutdown The op- rator must then wait almost 2 days for the Xe^{135} o disappear by radioactive decay At very high luxes this effect becomes so severe that even a mall temporary reduction in power may lead to ultimate subcriticality of the reactor

Another effect caused by Xe^{135} in high flux re- ictors is that if the reactor is large enough the Xe^{135} may force the power pattern into oscilla tions of 1 or 2 day periods Although this is not a serious dynamic problem it does emphasize the necessity of monitoring not only the total power but also the power pattern so that appropriate countermeasures may be taken

Reactor kinetics This is the study of the short term aspects of reactor dynamics with respect to stability safety against power excursion and de- sign of the control system Control is possible be cause there is a time lapse between successive fis- sions in a chain resulting from the finite velocity of the neutrons and the number of scattering and moderating events intervening and because a fraction of the neutrons is delayed See NEUTRON DELAYED

Prompt neutron lifetime This is the mean time between successive fissions in a chain and it is the basic quality which determines the time scale within which controlling effects must be operable if a reactivity excursion is touched off In thermal reactors the controlling feature is the time be- tween thermalization and capture because fast neutrons spend less time between collisions and the total time for moderation of a neutron is at most a few microseconds (μsec) Prompt neutron lifetimes vary from 10 to a few hundred μsec for light water reactors the shorter times being found in poorly reflected highly absorbing systems and the longer in well reflected systems in which time spent in the reflector is the dominating factor Other thermal reactors have longer lifetimes with some heavy water reactors having lifetimes as long as a few milliseconds Fast reactors have lifetimes of the order of 0 01–0 1 μsec, the controlling fac tor being the amount of scattering material used as diluent

Delayed neutrons Delayed neutrons are impor tant because a complete fission generation is not achieved until these neutrons have been emitted by their precursors A slightly supercritical reactor must wait until the delayed neutrons appear and il is delay allows time for the system to be brought under control The fraction of delayed neutrons β ranges from about $\frac{1}{3}$% for thermal fission of Pu^{239} and U [illegible] to about $\frac{3}{4}$% for thermal fission of U [illegible]35 to several per cent for some fast fission events In the [illegible] latter cases however the extra delayed neu trons have such short lifetimes that they are of only slight extra utility

In any case the influence of delayed neutrons is felt only to the extent that they are needed to main [illegible] When the system is supercritical [illegible]

tive sense the threshold between externally [illegible] trollable and uncontrollable excursions and it is for this reason that the dollar unit is popular in ex cursion analysis (a prompt critical system has a reactivity of 1 dollar)

Although it has now been established that a larger number of fission products emit delayed neu trons the distribution in time after fission of the delayed neutron emission rate is accurately repre- sented for all purposes by a sum of six negative ex ponentials For many purposes however a three, two or one group approximation is adequate

Reactors moderated by D_2O and Be have addi tional delayed neutrons contributed by photoneu tron reactions between the moderator and fission product γ rays Although not a large fraction this effect makes such reactors unresponsive to small reactivity fluctuations, and gives them unusual op- erational smoothness

Reactor period This is the asymptotic time re- quired for a reactor at constant reactivity to in crease its power by a factor e When a critical re- [illegible]

rise exponentially The period is the measure of this exponential rate

The relation between the reactor period and the reactivity is known as the inhour equation If l is the reactor lifetime in seconds β_i the fraction of delayed neutrons in group i λ_i the decay constant of group i delayed neutrons in sec^{-1} S reactor pe- riod in sec and ρ reactivity in thousands of milli kay

$$\rho = \frac{l/S + \sum_i [\beta_i/(1 + \lambda_i S)]}{1 - \sum_i [\beta_i/(1 + \lambda_i S)]}$$

is the inhour equation The equation has $I + 1$ so lutions of S for a given ρ I being the number of groups and there is always a real value of S with a higher value than the real part of any other solu tion This highest S is the period For very small values of ρ that is for very large periods this value is approximately

$$\left(l + \sum_i \frac{\beta_i}{\lambda}\right) \Big/ \rho$$

For very large ρ, and therefore small S the period is approximately $l/(\rho - \beta)$ This same result would be found if the delayed neutrons were thrown away completely

Reactivity coefficients There are several func tions relating changes in reactivity to changes in

able materials for research reactors in which flux peaks are useful for experimental purposes

The usual measurement of reflector worth is in terms of reflector savings defined as the difference in the reflected dimension between the actual core and one which would be critical without reflector Reflector savings are close to reflector dimensions for thin reflectors and approach an asymptotic value dependent upon core size and reflector con stitution as thickness increases

Reactor dynamics Reactor dynamics is con cerned with the temporal sequence of events when neutron flux power, or reactivity varies. The inclu sive term takes into account sequential events not necessarily concerned with nuclear processes, which may affect these parameters There are basi cally three ways in which a reactor may be affected so as to change reactivity A control element ab- sorbing rod or piece of fuel may be externally ac tuated to start up shut down or change reactivity or power level, depletion of fuel and poison buildup of neutron absorbing fission fragments and production of new fissionable material from the *fertile* isotopes $Th^{\prime 3}$ U^{34} U^{38} and Pu^{40} make reactivity depend upon the irradiation his- tory of the system and changes in power level may produce temperature changes in the system lead ing to thermal expansions and mechanical changes of its constituents with consequent change of re activity

Reactor control physics *Reactor control* phys ics is the study of the effect of control devices on reactivity and power level As such it includes a number of problems in reactor statics because the primary question is to determine the absorption of the control elements in competition with the other neutronic processes It is however a problem in dynamics given the above information to deter mine what motions of the control devices will lead to stable changes in reactor output

Particular problems occurring in the statics of *reactor control stem from the particular nature of* control devices Many control rods are so heavily absorbing for thermal neutrons that elegant refine- ments of neutron transport theory are needed to es timate their absorption Other types of control rods include isotopes with heavy resonance absorp tion and the interaction of such absorbers with U^{238} resonances must be examined The motion of control rods changes the material balance of reac tor regions and with water moderated reactors peaks in the fission rate occur near empty rod channels By virtue of high absorption of their con stituent isotopes some absorbing materials (for example cadmium and boron) burn out in the re- actor and a rod made of these materials loses ab- sorbing strength with time As a final example the motion of a control rod may change the shape of the power pattern in the reactor so as to bring sec ondary pseudostatic effects into play *See* REAC TOR NUCLEAR

Reactivity changes Long term reactivity changes may represent a limiting factor in the burning of nuclear fuel without costly reprocess- ing and refabrication As the chain reaction pro- ceeds the original fissionable material is depleted, and the system would become subcritical if some form of slow addition of reactivity were not avail- able *This is the function of* shim rods in a typical reactor The reactor is originally loaded with enough fuel to be critical with the rods completely inserted As the fuel burns out, the rods are with- drawn to compensate.

In order to decrease requirements on the shim system many devices to overcome reactivity loss may be used A burnable poison may be incorpo- rated in the system This is an absorbing isotope which will burn out at a rate comparable to or greater than the fuel Burnable poisons are there- fore limited to isotopes with very *high* effective neutron absorption cross sections Combinations of poisons, and the use of self shielding of poisons can in principle make the close compensation of considerable reactivity possible without major con- trol rod motions but the technological problems in their use are formidable

A more popular *method for compensating* reac- tivity losses is the incorporation of fertile isotopes into the fuel This is desirable because the neu- trons captured in the fertile material are not wasted but used to manufacture new fissionable material and also because (as with U^{38} in U^{-5} reactors or Pu^{40} in Pu^{39} systems) the fertile ma- *terial is normally found* mixed with the fissionable and isotopic separation may be circumvented or minimized Depending on the conversion ratio (new fissionable atoms formed per old fissionable atom burned) and the fission parameters *of the* materi- als a reactor so fueled loses reactivity relatively slowly and in some cases may show a temporary reactivity increase The various isotopes produced by successive neutron capture in uranium and plu- tonium must be considered in this problem the higher isotopes becoming prominent at very long *exposures*

A final consideration of long term reactivity is the extra parasitic absorption of the fission prod- ucts as formed At long exposure this absorption becomes significant because of the *relatively* high absorption of many of the fission products. At shorter exposures isotopes of very high absorp- tion mainly Sm^{149} and Xe^{135} are more promi- nent These materials have such high cross section- that they reach a steady state concentration rela tively quickly burning out by neutron capture a rapidly as they are formed in fission

Xe^{135} is particularly interesting because it cross section is abnormally large its fission yield i high it is preceded by an isotope of low cross sec tion and approximately 7 *hr half life* (I^{135}) and it undergoes β decay to the low absorption Cs^{35} with a half life of about 10 hr This combination of properties gives several interesting effects Chief *of these is that at high flux* a *reservoir of* I^{135} i formed which continues to decay to Xe^{35} even after the reactor is shut down Bec 135 I

else The term

levels of impurities The last two class ... ones usually used in the laboratory The American Chemical Society has established specifications and tests for purity for some chemicals. Materials which meet these specifications are labeled ... ACS Specifications

A special group of extremely pure chemicals are called primary standard' reagents These reagents are usually readily available easily purified and unreactive with components of air such as water

... volumetric analysis or for other purposes ... which impurities must be quite low in concentration

Care is necessary to prevent contamination by dust or by other chemicals Transfers should be made directly from the container by pouring and not with spatulas or other tools. No material should be returned to the container The maintenance of purity in an opened container is a problem

Selective and specific reagents Chemical reagents are often classified on the basis of utility

... ever there are some reagents which react with only a limited number of other chemicals These reagents are called selective reagents The silver salts of chloride bromide iodide thiocyanate, and of a few other ions are insoluble in water therefore silver nitrate is a selective precipitating reagent for these ions A limited number of reagents are known which react appreciably with only one ion under specified conditions These reagents are called specific reagents The specific property is determined by the product formed The ion dimensions the

... in the acidity of the solution and the presence of other ions can change the condition of the ion so that no reaction can occur Most inorganic reagents are at best selective Most specific reagents are organic in nature

These organic reagents have two types of reactive groups. One group forms electrovalent bonds by charge neutralization and the second group forms covalent bonds by sharing electrons The products are cyclic because both bonds are to the same ion, and they are called chelate compounds These chelates are frequently very insoluble in water and intensely colored making them very useful in analysis. [K.G.S]

Bibliography F Feigl *Specific and Special Reactions* 1940 American Chemical Society *Reagent Chemicals—ACS Specifications* 1956 J Rosin *Reagent Chemicals and Standards* 1955

Real variable

A variable whose range is a subset of the real numbers By extension the term is also used to refer to the theory of functions of one or more real variables This theory has to do with properties of broad classes of functions such as continuity, types of discontinuities differentiability of functions, oscillation and variation of functions and the various kinds of integrals See INTEGRATION

Real numbers Real numbers are those commonly used in the geometric theory of measurement. The integers and fractions, also called rational numbers are included among the real numbers In practice an irrational number x is specified by telling which rational numbers are less than x and which are greater than x Such a division of the rational numbers into two classes was used by J W R Dedekind as the formal definition of a real number and is called a Dedekind cut.

The system of real numbers has the familiar algebraic properties and is also ordered This order is related to the algebraic operations of addition and multiplication by the following properties

If $a < b$ then $a + c < b + c$ for every number c

If $a < b$ then $ac < bc$ for every number $c > 0$

If $a > 0$ and $b > 0$ there exists an integer n such that $a < nb$

...

open interval and denoted by $(a\ b)$ Here a and b are the ends of the interval but are not included in it. When the left end a is included, the notation is changed to $[a\ b)$ The interval $[a\ b]$ including both ends is called a closed interval A more complicated set, called the Cantor discontinuum, which is very frequently of use in the theory of functions may be constructed as follows From the interval $[0\ 1]$ remove the open intervals $(\frac{1}{3},\frac{2}{3})$, $(\frac{1}{9}\ \frac{2}{9})$ $(\frac{2}{3} + \frac{1}{9},\ \frac{2}{3} + \frac{2}{9})$ and in fact all intervals $(c - \frac{1}{3^n}\ c)$ where $c = d_1/3 + d_2/3^2 + \cdots + d_{n-1}/3^{n-1} + \frac{2}{3^n}$ and each $d_i = 0$ or 2 It may be proved that the remaining set contains just as many numbers as the original interval $[0\ 1]$ (see INFINITY) Another set to which reference will be made consists of all the rational numbers x satisfying the inequality $x^2 < 2$

An upper bound for a set S of numbers is a number U such that $x \leqq U$ for every x in S, and a lower bound is a number L such that $L \leqq x$ for every x in S A set need not have an upper bound for example, the set of all rational numbers A least upper bound for S is an upper bound U such that V is not an upper bound when $V < U$ There cannot be more than one least upper bound for a set of real numbers A distinguishing property of the real number system states that every set having an upper bound also has a least upper bound in the system. This property fails for the system consisting only of the rational numbers as is shown by the fact that the least upper bound of the set consisting of the rational numbers x such that $x^2 < 2$ is the irrational number $\sqrt{2}$

the physical state of the reactor The power coefficient is the change in reactivity per unit change in reactor power, the temperature coefficient relates reactivity to temperature change, and is often broken down into fuel, moderator, and coolant coefficients, for low power graphite reactors, there exists a barometric coefficient, one may define also coolant circulation rate coefficients and void coefficients

Because the reactivity is commonly a complicated function of all the pertinent variables, the reactivity coefficient generally is the coefficient of the first term in a series expansion of the reactivity about the operating point This in turn describes a linear theory of reactor dynamics The theory may be extended to reactivity effects of arbitrary type by considering reactivity coefficients as functionals

The basic problem of reactor dynamics is the specification of the power coefficient of reactivity The chain power affects reactivity which affects power, is thereby analyzable, using the power coefficient functional together with the reactor kinetic equations The power coefficient is, however predictable only in terms of changes in temperature and flow resulting from power changes in the system Thus the analysis of power coefficient implies exhaustive knowledge of system behavior

Reactivity coefficients may be prompt or delayed and most delayed effects can be characterized as either of decay or transport type. An example of the decay type of coefficient is the contribution of coolant temperature change to power coefficient Here a power pulse gives a thermal effect on the coolant which is instantaneously observable and which decreases exponentially with a time constant imposed by the heat transfer equations An example of a transport type of delay is the delay attributable to coolant circuit times Here, a finite time lapse exists between the cause and the observable response Effects due to fuel heating are examples of prompt effects

Some types of power coefficient yield dangerous or unstable situations. Thus a power coefficient may contain a prompt positive (autocatalytic) term and a larger delayed negative term Even though such a system may be stable against slow power level increases it will undergo a violent excursion whenever power is raised rapidly enough to outstrip delayed effects Again a system with prominent delayed effects of the transport type is always unstable beyond some critical power even if the effect opposes the power shift, here there is a possibility of phase instability

The dynamic behavior of a reactor is usually analyzed by techniques common to all feedback systems *See* SERVOMECHANISM [B I S]

Bibliography S Glasstone *Principles of Nuclear Reactor Engineering*, 1955, S Glasstone and M C Edlund *The Elements of Nuclear Reactor Theory* 1952, D J Littler and J F Raffle *An Introduction to Reactor Physics*, 1957, R L Murray, *Introduction to Nuclear Engineering*, 1954, R L Murray, *Nuclear Reactor Physics* 1957, *Reactor Physics Constants*, Argonne Natl. Lab., ANL-5800, 1958, H Soodak and E. C. Campbell *Elementary Pile Theory*, 1950, R. Stephenson, *Introduction to Nuclear Engineering* 2d ed., 1958 *Physics of Reactor Design*, vol. 5, United Nations, Proc Intern Conf Peaceful Uses Atomic Energy 1956, *Reactor Physics*, vol 12, *Reactor Physics and Economics*, vol 13, United Nations Proc. Second Intern Conf Peaceful Uses Atomic Energy 1958, US Atomic Energy Commission, *The Reactor Handbook Physics*, vol 1, AECD-3645, 1955 A M Weinberg and E. P Wigner, *The Physical Theory of Neutron Chain Reactions*, 1958.

Reagent chemicals

High purity chemicals used for analytical reactions, for the testing of new reactions where the effects of impurities are unknown, and, in general, for chemical work where impurities must either be absent or at known concentrations If the concentration of impurity in any reagent is critical, an analysis should be made

Methods of purification. Chemicals are purified by a variety of methods. The most common method is recrystallization from solution For many inorganic chemicals, a saturated solution is prepared in water at the boiling point. After filtration to remove insoluble matter, the chemical crystallizes out of solution as it cools The crystals are removed by filtration on a small scale or by centrifugation on a large scale, washed with water to remove impurity-containing solution on the surface, and dried. If the substance to be purified is not appreciably more soluble in hot solution than in cold solution, then one can use isothermal crystallization the removal of solvent at constant temperature by reducing the pressure The recrystallization process is not always successful as a purification method In some cases, precipitation of crystals from a saturated water solution by adding a second solvent for example ethyl alcohol, is the simplest method One difficulty is that the impurity crystals may have the same structure as the desired ones that is, the two crystals may be isomorphous. If water of hydration is present, storage in an atmosphere of known humidity may be necessary to obtain a definite hydrate Occasionally it is easier to remove the impurity by dissolving it in a solvent in which the desired material is not very soluble.

If the desired chemical is volatile and the impurities are not volatile sublimation is an effective method of purification For example iodine and ar

for example, the addition of pure ammonium carbonate solution to pure calcium chloride solution precipitates pure calcium carbonate.

Standards of purity. Commercial chemicals are available at several levels of purity Chemicals labeled 'technical' or 'commercial' are usually quite impure. The grade "USP" indicates that the chemical meets the

States Pharmacopeia and nothing else The term means only that the chemical is purer than ical ' Chemicals designated "reagent grade' alyzed reagent" are specially purified mate- which usually have been analyzed to establish vels of impurities The last two classes are es usually used in the laboratory The Amer Chemical Society has established specifica and tests for purity for some chemicals. Ma s which meet these specifications are labeled ts ACS Specifications '

active with components of air such as water carbon dioxide The total sum of impurities ld be less than 0 02% These reagents are used etermine the concentrations of solutions used olumetric analysis or for other purposes in ch impurities must be quite low in concentra

are is necessary to prevent contamination by t or by other chemicals. Transfers should be de directly from the container by pouring and with spatulas or other tools No material should returned to the container The maintenance of rity in an opened container is a problem

Selective and specific reagents Chemical re- ents are often classified on the basis of utility any chemicals are general reagents that is they ay react with many others For example any acid ll be neutralized to some extent by a base How er there are some reagents which react with only limited number of other chemicals These re- gents are called selective reagents The silver salts chloride bromide iodide, thiocyanate and of a w other ions are insoluble in water, therefore ilver nitrate is a selective precipitating reagent for hese ions. A limited number of reagents are known hich react appreciably with only one ion under pecified conditions These reagents are called spe- ific reagents The specific property is determined by the product formed The ion dimensions the charge densities and the electron arrangements of the reagent and the ion which react must fall within certain limits or no reaction will occur Variations in the acidity of the solution and the presence of other ions can change the condition of the ion so that no reaction can occur Most inorganic reagents are at best selective Most specific reagents are organic in nature

These organic reagents have two types of reac- tive groups One group forms electrovalent bonds by charge neutralization and the second group forms covalent bonds by sharing electrons The products are cyclic because both bonds are to the same ion and they are called chelate compounds These chelates are frequently very insoluble in wa ter and intensely colored making them very use ful in analysis [K C S]

Bibliography F Feigl, *Specific and Special Re actions*, 1940 American Chemical Society, *Rea- gent Chemicals—ACS Specifications* 1956, J Ro- *Reagent Chemicals and Standards*, 1955

Real variable

A variable whose range is a subset of the real num bers. By extension the term is also used to refer to the theory of functions of one or more real vari ables This theory has to do with properties of broad classes of functions, such as continuity, types of discontinuities, differentiability of func tions, oscillation and variation of functions, and the various kinds of integrals *See* INTEGRATION

Real numbers Real numbers are those com monly used in the geometric theory of measure- ment. The integers and fractions, also called rational numbers, are included among the real num bers In practice an irrational number x is specified by telling which rational numbers are less than x and which are greater than x Such a division of the rational numbers into two classes was used by J W R Dedekind as the formal definition of a real number and is called a Dedekind cut.

The system of real numbers has the familiar algebraic properties and is also ordered This order is related to the algebraic operations of addition and multiplication by the following properties

If $a < b$, then $a + c < b + c$ for every number c

If $a < b$, then $ac < bc$ for every number $c > 0$

If $a > 0$ and $b > 0$, there exists an integer n such that $a < nb$

Various subsets of the real numbers may be de- fined by means of inequalities The set of numbers x satisfying the inequalities $a < x < b$ is called an open interval and denoted by $(a\,b)$ Here a and b are the ends of the interval but are not included in it. When the left end a is included, the notation is changed to $[a\,b)$ The interval $[a\,b]$, including both ends, is called a closed interval A more complicated set, called the Cantor discontinuum which is very frequently of use in the theory of func tions, may be constructed as follows From the in terval $[0\,1]$ remove the open intervals $(\frac{1}{3}\,\frac{2}{3})$, $(\frac{1}{9}, \frac{2}{9})$ $(\frac{2}{3}+\frac{1}{9}\;\; \frac{2}{3}+\frac{2}{9})$, and in fact all intervals $(c - \frac{1}{3^n},\ c)$, where $c = d_1/3 + d_2/3^2 + \cdots + d_{n-1}/3^{n-1} + \frac{2}{3^n}$, and each $d_i = 0$ or 2 It may be proved that the remaining set contains just as many

inequality $x^2 < 2$

An upper bound for a set S of numbers is a num ber U such that $x \leqq U$ for every x in S and a lower bound is a number L such that $L \leqq x$ for every x in S A set need not have an upper bound, for exam ple, the set of all rational numbers A least upper bound for S is an upper bound U such that V is not an upper bound when $V < U$ There cannot be more

upper bound also has a least upper bound in the *system* This property fails for the system consisting only of the rational numbers as is shown by the fact that the least upper bound of the set consisting of the ra tional numbers x such that $x^2 < 2$ is the irrational number $\sqrt{2}$

the physical state of the reactor The power coefficient is the change in reactivity per unit change in reactor power, the temperature coefficient relates reactivity to temperature change, and is often broken down into fuel, moderator, and coolant coefficients, for low power graphite reactors, there exists a barometric coefficient, one may define also coolant circulation rate coefficients and void coefficients

Because the reactivity is commonly a complicated function of all the pertinent variables, the reactivity coefficient generally is the coefficient of the first term in a series expansion of the reactivity about the operating point. This in turn describes a linear theory of reactor dynamics The theory may be extended to reactivity effects of arbitrary type by considering reactivity coefficients as functionals

The basic problem of reactor dynamics is the specification of the power coefficient of reactivity The chain, power affects reactivity which affects power, is thereby analyzable, using the power coefficient functional together with the reactor kinetic equations The power coefficient is, however, predictable only in terms of changes in temperature and flow resulting from power changes in the system Thus the analysis of power coefficient implies exhaustive knowledge of system behavior

Reactivity coefficients may be prompt or delayed and most delayed effects can be characterized as either of decay or transport type An example of the decay type of coefficient is the contribution of coolant temperature change to power coefficient Here a power pulse gives a thermal effect on the coolant which is instantaneously observable, and which decreases exponentially with a time constant imposed by the heat transfer equations An example of a transport type of delay is the delay attributable to coolant circuit times Here a finite time lapse exists between the cause and the observable response Effects due to fuel heating are examples of prompt effects

Some types of power coefficient yield dangerous or unstable situations Thus a power coefficient may contain a prompt positive (autocatalytic) term and a larger delayed negative term Even though such a system may be stable against slow power level increases, it will undergo a violent excursion whenever power is raised rapidly enough to outstrip delayed effects Again a system with prominent delayed effects of the transport type is always unstable beyond some critical power even if the effect opposes the power shift, here there is a possibility of phase instability

The dynamic behavior of a reactor is usually analyzed by techniques common to all feedback systems See SERVOMECHANISM [B I.S]

Bibliography S Glasstone *Principles of Nuclear Reactor Engineering*, 1955, S Glasstone and M C Edlund *The Elements of Nuclear Reactor Theory*, 1952, D J Littler and J F Raffle *An Introduction to Reactor Physics*, 1957, R L Murray, *Introduction to Nuclear Engineering*, 1954, R L Murray, *Nuclear Reactor Physics*, 1957, *Reactor Physics Constants*, Argonne Natl. Lab., ANL-5800, 1958, H Soodak and E. C. Campbell, *Elementary Pile Theory*, 1950, R Stephenson, *Introduction to Nuclear Engineering* 2d ed., 1958 *Physics of Reactor Design*, vol 5 United Nations, Proc Intern Conf Peaceful Uses Atomic Energy 1956, *Reactor Physics*, vol 12, *Reactor Physics and Economics*, vol 13, United Nations Proc Second Intern Conf Peaceful Uses Atomic Energy 1958, US Atomic Energy Commission, *The Reactor Handbook Physics*, vol. 1, AECD 3645, 1955 A M Weinberg and E P Wigner, *The Physical Theory of Neutron Chain Reactions*, 1958.

Reagent chemicals

High purity chemicals used for analytical reactions, for the testing of new reactions where the effects of impurities are unknown, and in general, for chemical work where impurities must either be absent or at known concentrations If the concentration of impurity in any reagent is critical an analysis should be made

Methods of purification Chemicals are purified by a variety of methods The most common method is recrystallization from solution. For many inorganic chemicals, a saturated solution is prepared

filtration on a small scale or by centrifugation on a large scale, washed with water to remove impurity containing solution on the surface and dried If the substance to be purified is not appreciably more soluble in hot solution than in cold solution then one can use isothermal crystallization the re moval of solvent at constant temperature by reducing the pressure The recrystallization process is not always successful as a purification method

the same structure as the desired ones, that is, two crystals may be isomorphous If water of hydration is present storage in an atmosphere of known humidity may be necessary to obtain a definite hydrate Occasionally it is easier to remove the impurity by dissolving it in a solvent in which the desired material is not very soluble.

If the desired chemical is volatile and the impurities are not volatile sublimation is an effective method of purification For example, iodine and

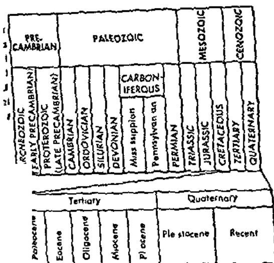
for example, the addition of pure ammonium carbonate solution to pure calcium chloride solution precipitates pure calcium carbonate

Standards of purity Commercial chemicals are available at several levels of purity Chemicals labeled 'technical" or commercial' are usually quite impure The grade USP" indicates that the chemical meets the [illegible] of

hine reamers for finishing remove metal by both nd and side cutting while the heavier rose reamers cut only on their chamfered ends See MACHINING OPERATIONS [A T]

Recent

A term roughly defining a sequence of geologic strata (the Recent Series) and also the time (late Quaternary) during which the strata were depos-



d Introduced by the British geologist Sir harles Lyell in 1833 the term is now widely con dered to represent the sediments postdating the e Age It is recognized however that the last eglaciation of middle and high latitudes has been rogressive, and various events have been sug ested as an arbitrary basis of separation By thers "recent" has been employed in an informal ense without specific definition In several Euro- pean countries the term Holocene is preferred to Recent but is used with essentially the same mean ing as Recent See QUATERNARY

Recent strata represent virtually every environ ment of deposition as they include all the sedi ments that are being deposited at present. Al though Recent time (in the formal sense) falls well within the range of C^{14} dating it lacks an ac cepted value for its duration because of uncertainty in fixing its inception It is held by some to repre sent a lapse of the order of 10 000 years The term hers in ited by 10 000 years as determined by C^{14} dating See PALYNOL OGY POSTGLACIAL VEGETATION AND CLIMATE RA DIOCARBON DATING [R F F]

Reciprocity, principle of

In the scientific sense a theory that expresses vari ous reciprocal relations for the behavior of some physical systems. Reciprocity applies to a physical system whose input and output can be interchanged without altering the response of the system to a given excitation Optical acoustical electrical and mechanical devices that operate equally well in either direction are reciprocal systems, whereas unidirectional devices violate reciprocity

f reciprocity facilitates the evalua

there is no need to consider any nonreciprocal com ponents when designing the system

Examples of reciprocal systems. Some systems that obey the reciprocity principle are (1) any electrical network composed of resistances, in ductances, capacitances, and ideal transformers (2) systems of antennas with restrictions given following Eq (2), (3) mechanical gear systems, and (4) light sources lenses, and reflectors

Devices that violate the theory of reciprocity are (1) transistors, (2) vacuum tubes, (3) gyrators, and (4) gyroscopic couplers Any system that con tains the above devices as components must also violate the reciprocity theory The gyrator differs from the transistor and vacuum tube in that it is linear and passive as opposed to the active and no

posed by Lord Rayleigh for acoustic systems and later generalized by J R Carson to include electro- magnetic systems Both mathematical expressions of the theory of reciprocity are closely related to the mathematical theorem known as Green's theo- rem (see GREEN'S THEOREM) The acoustical reci procity theorem of Lord Rayleigh is as follows In an acoustic system consisting of a fluid medium having boundary surfaces s_1 s_2 , s_k and sub- ject to no impressed body forces the surface in tegral

$$\int_s (p_1 v_{2n} - p_2 v_{1n})\, ds = 0 \qquad (1)$$

where p_1 and p_2 are the pressure fields produced respectively by the components of the fluid veloci ties v_{1n} and v_{2n} normal to the boundary surfaces s_1, s_2 s_k The integral is evaluated over all boundary surfaces

For a region containing only one simple source H L F Helmholtz has shown that the theorem can be expressed as follows a simple source at A produces the same sound pressure at B as would have been produced at A had the source been located at B In other words the response of a human ear at B due to a vibrating tuning fork at A is the same as the response of the ear at A due to the same tuning fork when located at B The human ear the tuning fork and the intervening acoustical media constitute a physical system that obeys the theory of reciprocity

Electromagnetic systems The generalization of Lord Rayleigh s theorem to electromagnetic sys tems can be mathematically expressed by the volume integral

$$\int_v \nabla \cdot (\mathbf{E}_1 \times \mathbf{H}_2 - \mathbf{E}_2 \times \mathbf{H}_1)\, dv = 0 \qquad (2)$$

Functions. Consider a function $f(x)$ defined on an interval $a \leqq x \leqq b$ By the oscillation of f on a subinterval $[c\ d]$ of $[a,b]$ is meant the number $\omega(c,d)$ = least upper bound of $|f(x_1) - f(x_2)|$ for x_1 and x_2 in $[c,d]$ The function f is continuous at a point x of $[a\ b]$ in case $\omega(c\ d)$ tends to zero when c and d approach x, and f is continuous on $[a\ b]$ when it is continuous at every point of $[a,b]$ A function f continuous on a closed interval $[a\ b]$ attains a greatest and a least value on $[a\ b]$ Moreover, it takes every value between $f(a)$ and $f(b)$ If f has a derivative f' on the interval $[a\ b]$, then f' also takes every value between each pair of its values, although f' may fail to be continuous However, a continuous function f may fail to have a derivative at any point whatever This suggests a useful classification of functions according to the number of continuous derivatives they possess

The variation of a function f on an interval $[a,b]$ is defined to be the least upper bound $V(a,b)$ of the sums

$$\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|$$

for all finite sets of points $x_0, \qquad , x_n$ satisfying $x_0 = a, x_{i-1} < x_i, x_n = b$ One can write $V(a,b) = +\infty$ when no upper bound exists and in all other cases one says that f has bounded variation A continuous function f may fail to have bounded variation, as for example

$$f(x) = x \sin (1/x) \qquad 0 < x \leqq 1$$
$$f(0) = 0$$

on the interval $[0\ 1]$ On the other hand a function of bounded variation may be discontinuous, however it always has right hand and left hand limits at each point, so all its discontinuities are jump discontinuities This follows readily from the fact that a function of bounded variation always is equal to the difference of two nondecreasing functions A very remarkable theorem states that every function f of bounded variation has a derivative $f'(x)$ except at the points of a set which can be enclosed in a sequence of intervals the sum of whose lengths is arbitrarily small

If S is a set composed of nonoverlapping intervals $[c_k\ d_k]$ the variation $V(S)$ of f over S may be defined as the sum of the variations $V(c_k\ d_k)$ If $V(S)$ tends to zero with the sum of the lengths of the intervals the function f is said to be absolutely continuous on $[a\ b]$ An absolutely continuous function f admits a generalization of the fundamental theorem of integral calculus namely

$$f(b) - f(a) = \int_a^b f'(x)\ dx$$

where the integral is a Lebesgue integral

Continuous functions may be characterized in a way different from the definition given above Thus a function f defined on a finite closed interval $[a\ b]$ is continuous on $[a,b]$ if and only if for every positive number ϵ there exists a polynomial function $p(x)$ which approximates $f(x)$ with an error less than ϵ on the entire interval $[a\ b]$ that is, $|f(x) - p(x)| < \epsilon$ for $a \leqq x \leqq b$ A variation of this condition (which may seem minor) yields a much more restricted class of functions than the continuous functions—that is, if one requires the approximating polynomials to be of the form

$$p(x) = \sum_{k=0}^{n} c_k(x - x_0)^k$$

where the coefficients c_k belong to a fixed infinite sequence $c_0, c_1, c_2 \qquad$, and x_0 is a point of $[a,b]$, the function f will have continuous derivatives of all orders and will have many other properties which are best understood when the variable x is allowed to be complex (*see* COMPLEX NUMBERS AND COMPLEX VARIABLES) On the other hand, if it is required only that there exist a sequence of polynomials p_n such that

$$\lim_{n \to \infty} p_n(x) = f(x)$$

for each x on $[a,b]$, then $f(x)$ may be discontinuous Such a function f is said to belong to Baire's

[illegible]

tion of classes of discontinuous functions.

[L.M.G.]

Bibliography L. M Graves, *The Theory of Functions of Real Variables*, 2d ed , 1956

Realgar

A mineral having composition AsS and crystallizing in the monoclinic system Realgar is found in short, vertically striated crystals, but more frequently is granular and in crusts. There is one pinacoidal cleavage the hardness is 1.5–2 (Mohs scale) and the specific gravity is 3 48 The luster is resinous and the color red to orange. Realgar is found in ores of lead silver, and gold associated with orpiment and stibnite It occurs with the silver and lead ores in Hungary, Czechoslovakia and Germany Good crystals have come from Binnenthal, Switzerland and Allchar, Macedonia In the United States it is found at Manhattan Nevada, Mercer Utah, and as deposits from geyser waters in Yellowstone National Park *See* ARSENIC, ORPIMENT STIBNITE [C.S.HL.]

Reamer

A multiple-cutting edge tool designed to enlarge or accurately size and finish an existing hole in solid material by removal of a small amount of the stock The cutting edges may be ground on the apexes between longitudinal flutes or grooves, or cutting may take place on chamfered edges at the end of the reamer

Reaming is performed either manually or by machine. Hand reamers may be gripped with a wrench on their square-ended tangs while machine reamer shanks are made either tapered or round to adapt to machine tools Hand reamers are fluted with a slight end taper to aid in starting Fluted ma

...inals 1 1 and E is measured *at terminals* then G_{21} represents a response *measured at* for an excitation at 2 2. *Mathematically the* eral behavior of the box to excitations at both of terminals can be expressed as

$$E_1 = G_{11}E_1 + G_{12}E_2 \qquad (8)$$
$$E_2 = G_{21}E_1 + G_{22}E_2$$

long as the response bears a linear relation to excitation

If in addition to its linear characteristic the stem satisfies the relation

$$G_{12} = G_{21} \qquad (9)$$

principle of reciprocity is obeyed and the vice will operate equally in either direction henever $G_{12} \neq G_{21}$ the system violates the theory reciprocity with the result that the response in e direction is different from that obtained in the her direction *See* NETWORK THEORY ELECTRICAL. [H S LA]

Bibliography D E Gray (ed) *American Institute of Physics Handbook* 1957 R F Harring on *Introduction to Electromagnetic Engineering* 958 International Telephone and Telegraph Corporation *Reference Data for Radio Engineers*, 1943 J D Krauss *Antennas* 1950 J A Stratton *Electromagnetic Theory* 1941

Recombination, genetic

The formation of new combinations of genes by the replacement of a portion of the genetic material from one cell lineage with its counterpart derived from another lineage Genetic recombination is a normal consequence of sexual reproduction, and the increased adaptability conferred on organisms by recombination is believed to be an evolutionary advantage that is responsible for the widespread occurrence of sex in diverse forms of life Genetic recombination is also known to occur by processes such as transduction in microorganisms that more nearly resemble infection than sexual union Recombination has provided geneticists with their most powerful experimental tool for investigating chromosome behavior and structure and for re solving the hereditary material into its genic elements *See* EVOLUTION ORGANIC

Genetic recombination requires for its detection the existence of inherited differences between organisms Such differences are due in most cases to physical differences localized in the linear chromosomes of the cell nucleus *See* CHROMOSOME THEORY OF HEREDITY

Two cells from separate lineages are brought together in sexual organisms by fertilization followed by fusion of their haploid nuclei each containing a single set of chromosomes to form a single diploid nucleus which contains two sets (*see* SYNGAMY) *Reassortment of genes* and chromosomes of the diploid into new combinations occurs when single sets of chromosomes are segregated into the haploid eggs sperm, or spores that result from the cell divisions during meiosis (*see* MEIOSIS)

An inherited difference referable to corresponding places (loci) in homologous chromosomes is said to be due to homologous that is allelic genes. These are conventionally represented by different forms of the same base symbol such as A versus a When an allelic difference is present the corresponding locus is said to be genetically marked Genes at nonhomologous loci are termed nonallelic and are symbolized by different base symbols, such as A or a at one locus versus B or b at a different locus Genes are known to be at different loci only if they are capable of recombining with one another See GENE

Effect of independent assortment. *Members of* chromosome pairs are normally assorted into the meiotic products at random with respect to the parentage and gene content of the nonhomologous chromosomes that accompany them In the case of two gene differences A, a and B b located in independently segregating nonhomologous chromosomes recombinations are therefore just as numerous as parental combinations among the haploid products of meiosis The combinations AB ab Ab and aB are equally frequent regardless of whether the diploid parent AB/ab was formed from the haploids AB and ab or from the haploids Ab and aB (Fig 1) *See* MENDELISM

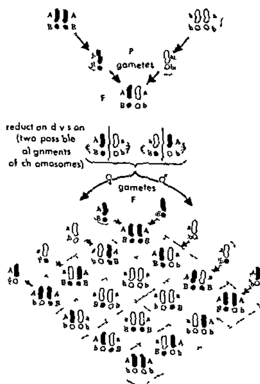


Fig 1 Diagram showing the combinations of *two* gene differences Aa and Bb located in independently segregating chromosomes *(From E W Sinnott L C Dunn T Dobzhansky Principles of Genetics 5th ed McGraw-Hill 1958)*

where E_1, H_1 are the electric and magnetic field vectors describing a state due to one electromagnetic source and E_2, H_2 describe another state due to a second source The above relation is valid as long as the medium is isotropic and the field vectors are finite, continuous, and vary according to a linear law (thus excluding ferromagnetic materials, electronic space charges, and ionized gas phenomena)

By means of Maxwell's equations, the relation of Eq (2) *can be expressed in another form when* restricted to systems of conduction current only

$$\int_v (\mathbf{E}_1 \cdot \mathbf{J}_2 - \mathbf{E}_2 \cdot \mathbf{J}_1)\, dv = 0 \qquad (3)$$

where J_1 and J_2 are the conduction current densities in an electromagnetic system respectively due *to the action of the external electric fields* E_1 *and* E_2

Equation (3) is readily applied to antennas and radiation If in Fig 1, J_1 is the resulting current density in antenna B due to an electric field E_1 established by antenna A, and J_2 is the current density in antenna A due to electric field E_2 established by antenna B, then $J_1 = J_2$ provided $E_1 = E_2$ The two emfs need not be applied at the same instant of time The integral in Eq (3) over all space reduces to an integral over the two antennas since J_1 and J_2 are zero elsewhere From this particular application of the reciprocity theorem it is seen that the transmitting and receiving patterns of an antenna are the same

The expression in Eq (3), when evaluated over *an N mesh electrical network, reduces to*

$$\sum_{j=1}^{N} V_{aj} I_{bj} = \sum_{j=1}^{N} V_{bj} I_{aj} \qquad (4)$$

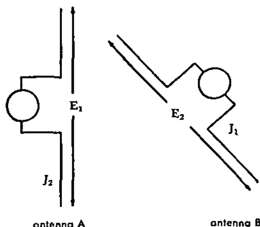


Fig 1 Antenna system

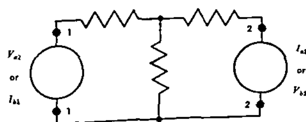


Fig 2 Two-mesh network

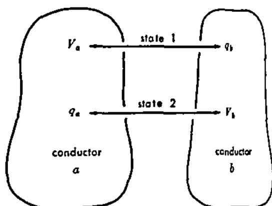


Fig 3 Charged conducting bodies

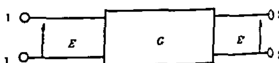


Fig 4 Four terminal, black box network.

where a and b are two different states of the network and the j subscript denotes in which of the N meshes the voltage and current are measured. For the two-mesh network in Fig 2, Eq (4) gives

$$V_{a1} I_{b1} = V_{b2} I_{a2} \qquad (5)$$

In words, if an emf source of magnitude V and zero internal impedance, when applied to terminals 1 1 produces a current I at terminals 2 2, then the same current I will be measured at terminals 1 1 when the emf V is applied *to terminals* 2 2 The preceding statement and Eq (5) are probably the most familiar form of the theorem of reciprocity

Electrostatic systems. The statement of reciprocity for electrostatics is

$$\int_v \rho_1 V_2\, dv = \int_v \rho_2 V_1\, dv \qquad (6)$$

where V_1 and V_2 are the electric potentials produced at some arbitrary point respectively due to the volume charge distributions ρ_1 and ρ_2 The integral expression in Eq (6) when applied to *the electrostatic system of two* charged conductors in Fig 3 becomes

$$V_a q_a = V_b q_b \qquad (7)$$

V_a *is the potential on conductor a* due to charge q_b on conductor b, and the remaining quantities are similarly defined In other words if a charge q on conductor b raises the potential of conductor a to V, *then the same charge on conductor a* raises the potential of conductor b to V

Electrical networks A somewhat different approach to reciprocity is the so-called 'black box,' *or two terminal pair method illustrated in* Fig 4. The box might contain a mechanical acoustic, optical, or electrical system The applied excitation or cause is E and the response or effect is E' The *ratio of E/E' (or E'/E) is the transfer function G* for the system within the black box Using the subscript notation of G_{12} when E is impressed at

that undergo segregation which may be either AB/ab which is the cis or coupling phase or Ab/aB the trans or repulsion phase This provides a basis for placing all genes of the same linkage group into definite positions with respect to one another on a linear genetic map (Fig 2) Map distance between two loci is expressed precisely as the mean number per 100 meiotic products of crossover or break points, that is crossovers in the marked interval The length of a map interval corresponding to 1% of crossovers is usually called one map unit although special terms such as morgan or centimorgan have been proposed to define such a unit

Crossover probabilities are additive whereas recombination values are not Crossovers and recombinations are equal for short intervals but recombinations across long intervals are less frequent than crossovers within them because of the occurrence of double and other multiple crossovers The most accurate maps are therefore based on recombination values for short intervals Although linkage groups are known whose length exceeds 200 map units recombination of any two genes does not normally exceed 50% because of multiple crossovers and the fact that each exchange involves only two out of four chromosome strands

Relation of genetic and cytological maps Genetic linkage groups can be assigned to particular chromosomes on the basis of parallelisms in the distribution through meiotic segregation and fertilization of genetic markers and of microscopically recognizable chromosome differences Sex linked genes and the recognizably different sex chromosomes provide the most common opportunity for accomplishing this Other linkage groups can be assigned to their respective chromosomes by using structural rearrangements heteromorphic chromosomes characterized by differential features such as knobs or atypical numbers of

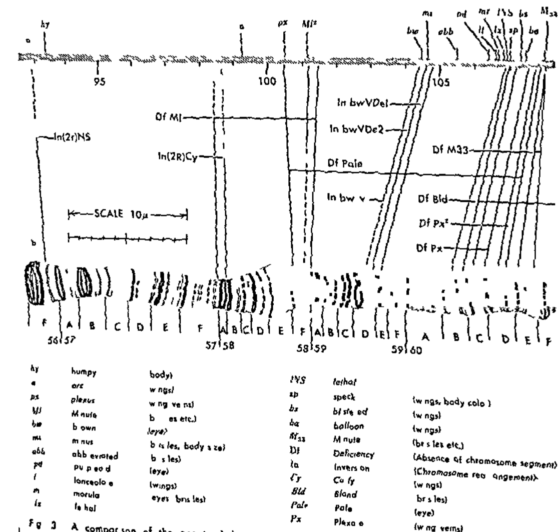


Fig 3 A comparison of the genetic linkage map (a) and corresponding section of the salivary-gland chromosome (b) The region of the chromosome shown is the end portion of the right limb of the second chromosome of Drosophila melanogaster (After C B Bridges from E W Sinnott L C Dunn T Dobzhansky Principles of Genetics 5th ed McGraw Hill 1958)

Effects of linkage and crossing over. Differences at different gene loci within a pair of homologous chromosomes do not ordinarily behave independently of one another during meiosis but remain in their original combinations unless a mutual exchange of equivalent segments occurs between paired homologs through a process known as crossing over Genes are known to be linked that is, to be located in the same or homologous chromosomes when parental gene combinations occur more frequently than recombinations among the products of meiosis The combinations AB and ab are produced more frequently than Ab and aB if the diploid parent came from $AB + ab \rightarrow AB/ab$, less frequently if the parent came from $Ab + aB \rightarrow Ab/aB$

Linkage in bacterial transductions and transformations is indicated when two inherited differences are transferred simultaneously from donor into recipient with a higher frequency than would be expected from chance coincidence *See* BACTERIAL GENETICS

Genetic linkage connotes mechanical relationships between genes during reassortment because of their being in homologous chromosomes It does not imply any functional similarity among the genes that are linked with respect to their effects on the organism However related functions are found together in some cases of very closely linked recombinable units (*see* PSEUDOALLELES) Also, certain genes concerned with sequential biosynthetic reactions are closely linked in bacteria.

Linkage groups All the gene loci that show linkage with one another make up a single linkage group and when enough genes have been tested the number of linkage groups is expected to equal the haploid chromosome number of the species. This has been demonstrated for various species of *Drosophila* fruit flies with 3–6 linkage groups for corresponding chromosome numbers in different species and also for maize with 10 barley with the fungus *Neurospora* with 7 the fungus *Aspergillus* with 8 the chicken with 6 and the garden pea with 7 linkage groups

Not all linkage groups are mapped in other organisms such as the house mouse with 16 groups known for 22 chromosomes the silkworm *Bombyx* with 15 groups for 28 chromosomes the Japanese morning glory *Pharbitis* with 10 groups for 15 chromosomes and the tomato with 11 groups for 12 chromosomes The bacterium *Escherichia coli* and certain bacterial viruses appear to possess only one linkage group Knowledge of linkage in man is rudimentary

Genetic maps The strength of linkage between genes at two specific loci depends upon their distance apart but is independent of their steric arrangement in the two homologous chromosomes

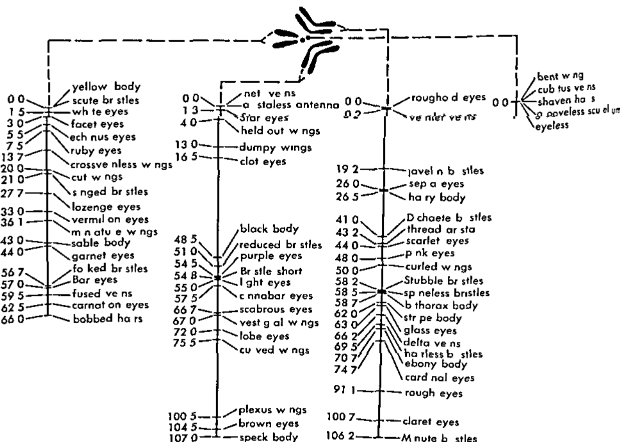


Fig 2 A genetic or linkage map of the four chromosomes of *Drosophila melanogaster* showing the relative positions of some of the more important genes Figures refer to distances from the upper end of the chromosome as determined from the percentages of recombination observed in linkage experiments (From E W Sinnott L C Dunn T Dobzhansky *Principles of Genetics* 5th ed McGraw Hill 1958)

Interference If the occurrence of two or more ossovers within neighboring regions is less fre uent than would be expected by chance coinci ence *interference* or as it is commonly termed hiasma interference is said to be *positive* Inter rence between two marked regions may be ex ressed quantitatively in terms of coincidence which is the ratio of the observed frequency of dou le crossovers to the frequency expected if the two vents were independent In most organisms but not in the homothallic fungus *Aspergillus* inter ference is strong or complete between short neigh boring regions comprising 10 map units or less and decreases for increasingly longer intervals Some cases of apparent multiple crossovers in very short regions may be due to nonreciprocal gene conversion which is discussed later in this article Inferences about interference based on data from random meiotic products depend upon assumptions concerning how the four strands of a tetrad are in volved in successive exchanges

S nd relat ons The fact that each crossing

onrandom participation of strands which is ermed chromatid interference would have impor

luites u c v

from crosses having three or more marked loci Al hough data from a number of experiments are con sistent with random involvement substantial evi dence from fungi and from flies suggests that the same two strands are more likely than not to be in volved in successive exchanges especially in short ne ghboring regions Interpretation of these results depends upon the unsettled question of whether cross ng over occurs between sister strands

Meiotic exchanges between sister strands which are identical replicates would not result in recom b nation of marker genes but might neverth eless have indirect effects on interference or on strand relations in multiple exchange tetrads Labeling with radioactive isotopes has enabled sister strand

at sister strand exchanges do not occur during iosis in such a way as to participate in interfer ce on an equal basis with genetically detectable changes See MITOSIS

Cytological observations Cytological observa ons have been too gross to contribute much di ect information and the microscope has served n ely to reveal later consequences of crossing ver to confirm inferences from more precise ge etic data and to explain anomalies of genetic re ombination in terms of chromosome structure or umber Good evidence exists that each genetically ffective exchange results in a chiasma which can e observed under favorable conditions as a cross shaped interchange between two members of a tetrad of chromosome strands. However chiasma data must be interpreted cautiously because in less favorable material structures indistinguishable from chiasmata may occur in the absence of ge netic exchange

Mitotic crossing over In sexual organisms re combination normally occurs with a high fre quency only at meiosis However both in insects and in fungi crossing over has been shown to occur with low frequencies in diploid cells that subse quently continue to divide mitotically as diploids In both these cases, reciprocal exchange occurs be tween nonsister chromatids at a 4-strand stage similar to meiotic crossing over

Delimitation of individual genes The individ ual gene has been variously defined as the small est unit in a linear array capable of undergoing re combination with other such units (the recon) or capable of undergoing detectable mutational change (the muton) or capable of functioning normally only if nondefective subunits lie in the

[illegible] han in separate homo-

[illegible]

nitions See GENE ACTION

Gene conversion Nonreciprocal recombination sometimes occurs within a short region so that one of a pair of segregating genetic units is repre sented in more than two of the four products com prising a single meiotic tetrad or a new combina tion of genic subunits appears that is not balanced b[illegible] reciprocal recombinant These exceptions to

[illegible]

A direct demonstration of gene conversion re quires that tetrads of meiotic segregants be ana lyzed Evidence from random meiotic products is necessarily indirect the occurrence of nonmutant progeny from crosses between two physiologically similar mutants has been interpreted in terms of gene conversion in a number of cases where neigh boring gene markers on either side of the mutant region are not recombined as frequently among the new types as would be expected if the nonmutants had arisen from simple crossing over between parts of a complex locus

Critical direct evidence for conversion is con fined to a few cases in fungi where specific genes undergo conversion in fewer than 1% of meioses at most Indirect evidence in fungi insects and seed plants suggests a more general occurrence Gene conversions are correlated with the occurrence of normal complementary crossing over at or near the locus involved The underlying mechanism is un known Hypotheses have considered clustered mul tiple exchanges within a complex locus, excep tional modes of gene replication or mutational changes involving extragenic controlling elements Failure to recognize conversion as such could lead to spurious conclusions regarding interference and regard ng the degree of linear resolution achieved in recombination experiments

homologs of a particular chromosome such as the trisomic or monosomic condition

Localization of specific genes within particular chromosome regions may be accomplished by using deficient chromosomes where the absence of a certain recognizable region permits the expression of whatever recessive genes are carried in the corresponding region of a nondeficient homologous chromosome This method is subject to possible errors resulting from position effects A more general method for assigning loci of the genetic maps to specific chromosome regions depends on the existence of heteromorphic chromosome features or rearrangements, and makes use of crosses when both genetic and cytological markers are segregating simultaneously

Gene sequences are identical on genetic and on cytological maps, but relative lengths of intervals between corresponding loci may differ (Fig 3) Rearrangements in chromosome structure such as inversions and translocations are reflected in changed linkage relations *See* CHROMOSOME ABERRATION

Tetrad analysis In most organisms, including those higher plants and animals that are best known genetically, only a random sample of meiotic products consisting of spores, eggs, or sperm that have originated from many separate meiotic segregations can be subjected to genetic analysis In some organisms, notably lower plants such as fungi, algae, and bryophytes, the four products of an individual meiosis can be recovered for analysis as a group of four spores (quartet or tetrad) Each member of such a spore tetrad contains one of four chromosomes from the tetrad of strands within which crossing over occurred

Genetic analysis of tetrads provides more direct and complete information regarding recombination mechanisms than it is possible to obtain from strands collected at random especially information regarding the reciprocal nature of exchanges, their occurrence at the 4-strand stage, and the relations between strands involved in multiple exchanges Tetrad analysis also enables linkage relations between centromeres and gene markers to be established

Mechanism of crossing over. Genetically detected crossing over entails actual physical exchange of microscopically distinguishable chromosome segments Equal, homologous segments that extend from one point of crossing over to another, or to the end of the chromosome, are exchanged, two complementary recombinant chromosome strands hence result from a single exchange (Fig 4) Individual exchanges affect only two of the four homologous chromosome strands that are recovered in the four products from an individual meiosis Crossing over therefore occurs between individual strands at or after the time when paired chromosomes split to form a four stranded structure containing two identical sister strands from each homolog In summary, when a single exchange occurs between loci *c* and *d* in a diploid containing

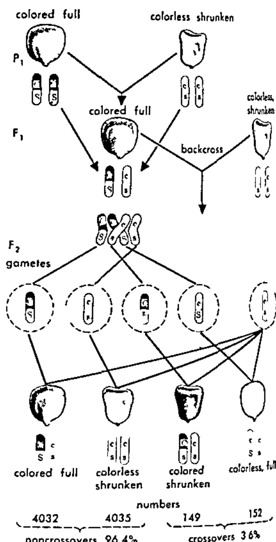


Fig 4 The chromosome interpretation of linkage and crossing over, illustrated by the behavior of the genes for colorless aleurone and for full or shrunken endosperm in corn (*From E W Sinnott L C Dunn T Dobzhansky, Principles of Genetics 5th ed, McGraw-Hill, 1958*)

chromosomes *ABCDEF* and *abcdef* the four meiotic products that result are *ABCDEF*, *ABCdef* *abcDEF* and *abcdef*

It is not known whether each primary crossing over event involves breaks and rejoinings of preformed structures or whether it is a consequence of replication along paired homologous chromosomes that serve as alternate templates

Crossing over frequencies within marked regions are not absolute but are capable of wide variations depending on genetic constitution, chromosome some pair The distribution of exchanges within and among pairs of homologous chromosomes is not random, occurrence of one exchange generally decreases the probability of another occurring nearby This is known as interference When multiple crossing over occurs, a second exchange may involve the same two strands as the first or only one of them or neither

on current sensitive paper, or a light beam m or light sensitive paper which can be con usly developed during use

ect acting instruments. Figure 1 shows a le direct acting, circular chart, pressure re- ng instrument Direct action operation is suit- when the primary detector has sufficient torque ercome the frictional loads of the bearings and ing means

ultiple record, direct acting, circular chart re ers are made with as many as four independ neasuring systems The pen arms are arranged nat they may pass each other without interfer The marking means may consist of an inte- assembly of inkwell and writing tip or a writ tip connected by capillary tubing to a ionary ink reservoir Figure 2 illustrates a four recorder

Multiple recording may also be made with a e type chart With this type the record of each asured quantity is restricted to a limited section he chart Simultaneous recordings of the phase tages of a polyphase electrical system are com nly made this way

A direct acting strip chart recording mechanism shown in Fig 3 Instruments of this type are de for use with any one of several nominal chart lths from 3 to 6 in Generally the chart moves tically downward with time Some recording in uments with narrow charts employ a horizontal ve to the left with the marking means moving ward for increasing values of the measured antity This arrangement provides first quadrant esentation commonly used in manual plotting The circular chart is usually driven by a small nchronous motor or spring wound clock Chart eeds range from one revolution in 15 min to one volution in 30 days Strip chart instruments gen ally employ an electric motor drive Chart ngths from 90 to 120 ft are standard Chart peeds from ½ in /hour to 6 in /sec can be pro

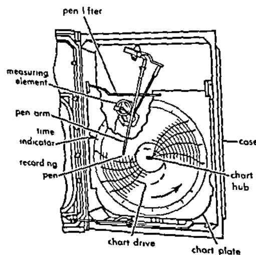


Fig 1 Single-pen, direct-acting circular-chart recorder (From D M Considine ed Process Instruments and Controls Handbook, McGraw Hill 1957)

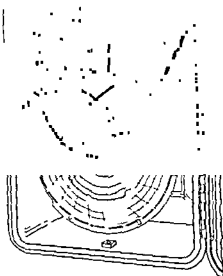


Fig 2 Four pen, direct-acting, circular-chart recorder (From D M Considine, ed, Process Instruments and Controls Handbook, McGraw Hill, 1957)

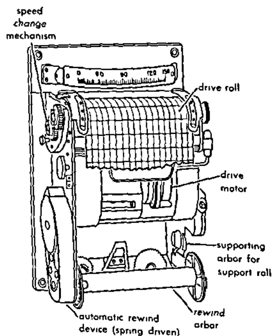


Fig 3 Strip-chart drive assembly (From D M Considine, ed, Process Instruments and Controls Handbook McGraw Hill, 1957)

vided Direct acting recording instruments are suit able for many measurements, such as temperature, pressure, flow, level, electric current, voltage, and power

A modified form of direct acting recording in strument uses an auxiliary source of power for op eration of the marking means This arrangement is employed when the measurement energy level is too low to overcome the frictional load of the marking means In this instrument the stylus does not

Bacteria and viruses. The mechanism of crossing over between homologous chromosomes seems to be fundamentally similar in all organisms above the bacteria, and normally to involve reciprocal exchange occurring at the 4-strand stage Genetic recombination in bacteria and bacterial viruses (bacteriophages) is overtly dissimilar from that in higher organisms in several respects, of which the most striking is that reciprocal products of recombination are not recovered from individual recombinational events This may reflect a fundamental difference in mechanism or it might in some cases reflect postrecombinational elimination of one of the reciprocal products perhaps because it is part of an incomplete or inviable complement Only a portion of the genetic complement from the donor cell lineage usually participates in bacterial recombination whether this be preceded by cell contact as in sexual union by transfer of genetic material by a bacteriophage vector, as in transduction, or uptake by the cells of molecular deoxyribonucleic acid (DNA) as in transformation

In bacteriophages recombination probably results from formation of a replica alternately on two different templates rather than from breakage and reunion of preformed structures Double crossovers in short intervals coincide with a frequency greater than chance Genetic information is contained completely in the bacteriophage DNA *See* BACTERIOPHAGE

Molecular basis. Evidence of many kinds indicates that DNA plays a fundamental genetic role for which its complementary duplex aperiodic linear structure seems uniquely appropriate Prospects of relating crossing over on the chromosomal level to replication and recombination on a molecular level have been increased by the demonstration (using radioactive isotopes) that a simple relation exists between DNA replication and the duplication of plant chromosomes undergoing mitosis Old and new daughter strands are clearly distinguishable when chromosomes have duplicated twice following removal from the source of label Experiments by J H Taylor (1958) have shown that labeled DNA is retained intact along the entire length of one strand whereas its sister remains unlabeled except when complementary sister strand crossovers are seen to have occurred Similar experiments in bacteria and in bacterial viruses indicate that a labeled DNA template is conserved substantially intact when new replicates are made *See* DEOXYRIBONUCLEIC ACID, GENETICS [D D P]

Bibliography *Exchange of Genetic Material Mechanisms and Consequences*, Cold Spring Harbor Symp Quant Biol, vol 23, 1958, K Mather Crossing over *Biol Revs*, 13 252 1938, W D McElroy and B Glass (eds), *A Symposium on the Chemical Basis of Heredity*, 1957, H J Muller, *Genetics, Medicine and Man*, 1947, G Pontecorvo *Trends in Genetic Analysis*, 1958, E W Sinnott, L C Dunn and T Dobzhansky, *Principles of Genetics*, 5th ed, 1958, C P Swanson, *Cytology and Cytogenetics*, 1957, *Symposium on Genetic Recombination*, J Cellular Comp Physiol vol 45, suppl 2, 1955, J H Taylor, Sister Chromatid exchanges in tritium labeled chromosomes, Genetics, 43 515, 1958

Recording

Any process for preserving signals, sounds, data or other information for future reference or reproduction Processes in common use include magnetic tape and wire recording, disk recording as on phonograph records, photographic recording of wave forms, and facsimile recording of photographs and other material Less commonly used are electromechanical, electrothermal, electrochemical electrolytic, carbon pressure and other recording processes The term recording is also used to denote the end product of a recording process, such as the magnetic tape, disk recording or paper chart from a graphic level recorder

Common speeds used for disk recordings are 16⅔, 33⅓, 45 and 78 rpm Speeds for magnetic tape are lower or higher multiples of 30 in /sec such as 60, 15, 7½, and 3¾ in /sec *See* MAGNETIC RECORDING, *see also* DATA PROCESSING SYSTEMS DISK RECORDING, FACSIMILE, OPTICAL RECORDING RECORDING INSTRUMENTS, GRAPHIC, SOUND RECORDING, WIRE RECORDING [J M.]

Recording instruments, graphic

Instruments that make a graphic record of one or more quantities as a function of another variable, usually time Recording instruments may have any type of sensing device such as pressure temperature voltage or weight The sole identifying feature is that they make a graphic record of the quantity being measured An instrument with a *recording device* often uses the suffix graph such as a barograph or oscillograph in place of the suffix used with a similar nonrecording instrument, such as a barometer or oscilloscope

Recording instruments are classified according to their principle of operation They are either direct acting in which case the marking device is mechanically connected to and directly operated by the primary detector or indirect acting in which case the measurement energy of the primary detector is increased through some intermediate means to actuate the marking device These intermediate means are usually mechanical, electrical electronic or photoelectric

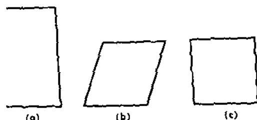
In addition to the primary classification recording instruments are classified by their exhibiting means recording means number of marking devices and marking means The exhibiting means are either circular charts or strip charts The recording means may be continuous intermittent (marking device retracted from chart between measurements) or sequential (more than one variable is intermittently recorded) The instrument may have one or more marking devices The marking means may be an inking pen on a paper chart, an inked typed impression on a paper chart a heated stylus on coated paper a mechanical stylus on coated or chemical ed paper an

on as the sum of two or more quantities,
root extraction, multiplication and division
recorder The function plotter, or X Y re-
provides a record of one variable against
er (not time) The instrument includes an in

:bli
iA)
:ts-
2
al Measuring Instruments, ASA C39 4, 1956,
M Considine (ed), *Process Instruments and*
trols Handbook, 1957

ctangle

rectangle is a parallelogram whose adjacent
les are perpendicular A rhombus is a plane
adrilateral having four equal sides. A square is



(a) Rectangle (b) Rhombus. (c) Square

a rectangle with four equal sides and is therefore also a rhombus The area of either the rectangle or the rhombus is equal to the product of the base times the altitude ($A = bh$) For a square of side b, the area A is equal to b times b, it is written b^2 and called b square" The numbers 0, 1, 4 9, 16, are called squares, since each has the form n^2 where n is an integer See QUADRILATERAL, SQUARE. [J S T]

Rectifier

A *nonlinear* circuit component that allows more current to flow in one direction than in the reverse direction An ideal rectifier would be one that allowed current to flow in one direction unimpeded but allowed no current to flow in the other direction. Thus, ideal rectification might be thought of as a switching action with the switch closed for current in *one direction* and the switch open for current in the *other direction*. Rectifiers are used primarily for conversion of alternating *current* (ac) to direct current (dc) However, rectification occurs in other types of circuits, such as detectors of amplitude-*modulated* waves, and limiters See POWER SUPPLY, ELECTRONIC.

A number of *rectifier* elements are in use. The vacuum tube rectifier is *used for* moderate power requirements Its resistance to current flow in one direction is essentially infinite, as the tube does not conduct when the plate is negative with respect to the cathode In the other (forward) direction, its resistance is small and almost constant (see DIODE, VACUUM) Gas tubes, used primarily for higher power requirements, also have a high resistance in the reverse direction (*see* GAS TUBE, MERCURY VAPOR RECTIFIER) The semiconductor, or metallic disk rectifier, has the advantage of not requiring a filament or heater supply This type of rectifier has approximately constant forward and reverse resistances, with the forward resistance being much smaller (*see* SEMICONDUCTOR RECTIFIER) Mechanical rectifiers can also be used The most common is the vibrator, but other devices are also used (*see* MECHANICAL RECTIFIER, VIBRATOR)

A rectifying element can be illustrated by assuming a device having a forward resistance R_1 and a reverse resistance R_2 much greater than R_1 A sinusoidal alternating voltage $E_m \sin 2\pi ft$ is applied to the rectifier, where E_m is the maximum value of the applied voltage, f is *the frequency of the voltage wave, and t is time* The magnitude of the current in the forward direction is $(E_m/R_1) \sin 2\pi ft$ This current flows from t equals 0 to $\frac{1}{2}f$, or for one-half the cycle of the alternating voltage wave. The average forward current averaged over one cycle, is $E_m/\pi R_1$ The reverse current has the magnitude $(E_m/R_2) \sin 2\pi ft$ and flows from t equals $\frac{1}{2}f$ to $1/f$, or for the other half cycle The average reverse current, averaged over one cycle, is $E_m/\pi R_2$ The net forward average current is $E_m(R_2 - R_1)/\pi R_1 R_2$

If the reverse resistance R_2 is extremely large compared to R_1, the average current approaches $E_m/\pi R_1$ If the average current is subtracted *from* the current flowing in the rectifier, an alter*nating* current results This ripple current flowing *through* a load produces a ripple voltage which is *often undesirable* Filter and regulator circuits are used to reduce it to as low a value as *is required* *See* FILTER ELECTRIC, VOLTAGE REGULATOR

Half wave rectifier circuit A half wave rectifier *circuit is* shown in Fig 1 The *rectifier tube* is a vacuum diode, which allows current to flow in the forward *direction from* the anode to the *filament* but allows practically no current to flow in the reverse *direction from* the filament to the anode The

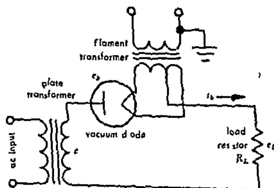


Fig 1 Half wave vacuum diode rectifier

normally touch the chart It is pressed against the chart periodically by an auxiliary motor mechanism By the use of a coated chart or an inked ribbon between the stylus and the chart, an intermittent record is made

Indirect acting instruments. Indirect acting recording instruments may employ mechanical, pneumatic, electrical, or electronic power amplification, the latter being the most common Feedback circuitry is generally used to eliminate effects of amplifier gain instability (*see* SERVOMECHANISM). Either of two types of feedback are used in electrical measuring recorders, current or position.

Figure 4 is a schematic diagram of a current-feedback circuit The voltage input (error signal) to the amplifier is the difference between the measured quantity e and the voltage drop IR across resistor R At balance only a small error signal is necessary to sustain the amplified feedback current I When the measured quantity changes, the error signal increases This causes the amplifier output current to change until the new value of feedback voltage IR differs from the new measured quantity by only a small amount The feedback current is connected in series with a direct deflecting recording milliammeter Because the energy level of the signal to the recorder is high, this instrument does not have the torque limitation of the direct acting electrical recording instrument This type of instrument is suitable for the measurement of rapidly varying quantities, with rates of change as high as 100 cps

Figure 5 is a schematic circuit diagram of a potentiometer type position feedback instrument The measured quantity e is determined by positioning a contact on a slidewire so that the voltage drop across the portion of the slidewire from one end to the contact is equal to e By maintaining a known fixed current in the slidewire, the position of the contact, as read on the associated chart, is a measure of the input quantity These measurements are made automatic by connecting the error signal to the input of an amplifier The amplified output drives a motor, which moves the contact until the error signal is effectively reduced to zero Instruments of this class usually use a converter and an ac amplifier and motor The converter changes the dc error signal to alternating current, which is amplified to operate the balancing motor This combination eliminates the need for dc amplification, which is difficult to stabilize

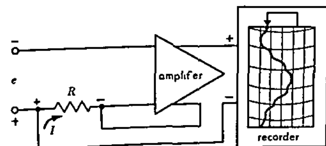


Fig 4 Schematic diagram of a current feedback circuit

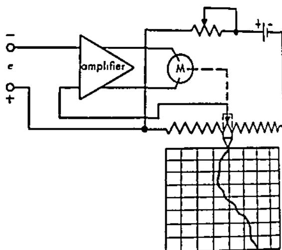


Fig 5 Schematic diagram of a position feedback circuit

Position feedback instruments are made with a wide variety of characteristics They may use circular or strip charts Nominal strip-chart widths are 3, 4, 6, 8, 10 and 12 in Full scale step-response time from ¼ to 30 sec can be provided Full scale range can be as low as 100 μv Nominal accuracy of this class of instrument is 0 25% of scale span. By special selection of components an accuracy of 0 1% is attainable

The use of a relatively high power balancing motor makes possible the operation of auxiliary apparatus from the output shaft of the motor drive. Alarm switches, electrical and pneumatic control mechanisms, integrators, and analog to-digital converters can be operated by the instrument without impairing its measuring characteristics Manual or ... speed change, and

viding continuous measurement of two independent variables Separate measuring circuits amplifiers, and balancing motors are used for making each record The pens, using different colors of ink may be arranged to pass each other, or they may be restricted to a limited portion of a zoned chart

A large number of variables can be recorded on a single chart by sequential means A motor-driven switch connects the input circuit sequentially to the primary detectors After each balancing operation a printing mechanism is pressed against the chart marking the value of the measurement The primary detector associated with each individual record is identified by a distinctive color or by a printed numeral or letter As many as 28 variables can be recorded on a single chart The time per point may be as short as 1 sec or as long as 1 min.

The position feedback recording instrument has many applications By the use of suitable transducers, such as thermocouples, strain gages, tachometers, and photocells, nonelectrical quantities can be recorded It is suitable for use with a variety of measuring circuits, such as potentiometers and dc and ac bridges Circuit combinations can be used to record the resul uous analo com

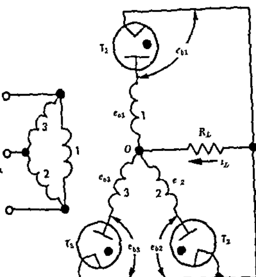


Fig 5 Three phase half wave rectifier

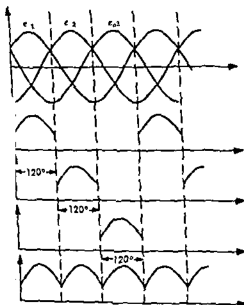


Fig 6 Transformer voltages tube currents and load current in a three-phase half wave rectifier

conduct and tube T_1 stops conducting The secondary voltages e_1 e_{s2} and e_{s3} are shown in Fig 6 and the tube *currents* are indicated as i_1 i_2 and i_3 The resulting load current i_L is *also* shown in Fig 6 This *current is* much closer to a true direct current than is the *current for the* single-phase circuits of Figs 1 *and* 3 The ripple voltage is much lower and less elaborate filter circuits *are* needed to smooth the output wave The tubes used in Fig 5 *might* also be ignitrons or pool cathode tank rectifiers

Another common polyphase rectifier circuit is the three phase *full* wave or six phase half wave ... of Fig 7 In this circuit the tubes conduct ... instead of 120° for the circuit of Fig 5

The ripple voltage for the full wave circuit is much smaller Many other polyphase rectifier circuits are possible

Bridge rectifier circuits Bridge rectifier circuits are useful in both single-phase and polyphase applications where a transformer must be used whose secondary has no center tap or where dc voltages approximately equal to the total secondary voltage of the transformer must be obtained Another use of the bridge circuit is in ac rectifier type meters The bridge circuit is shown in Fig 8 Four separate half wave rectifier tubes are used When the left hand side of the transformer secondary is positive current flows through tube T_1 the load resistor R_L and tube T_3 When the secondary voltage reverses the current flows through tubes T_2 and T_4 passing through R_L in the same direction as during the first half cycle

Parallel rectifiers If greater current is desired two or more rectifiers can be paralleled This is particularly simple if the tubes have the same ratings and are vacuum tubes For paralleling gas tubes small resistors are put in series with the

...

ing the primary voltage of the power transformer or by changing a resistance in series with *the load* resistor The first has the disadvantage of *being* expensive the second leads to poor *efficiency* A more convenient and less expensive method *is to* control the angle at which the rectifier tube *starts* to conduct Special gas tubes that accomplish this control are thyratrons ignitrons and excitrons. Thyratrons are hot cathode gas tubes with a large grid structure that prevents the arc from *being* ignited until the correct voltage is applied to the grid An ignitron is a cold cathode pool type tube with an igniter grid actuated by an electrical *pulse* The igniter of the ignitron requires a *substantial* amount of power which is usually supplied by an

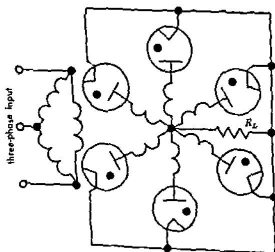


Fig 7 Three phase full wave or six phase half wave rectifier

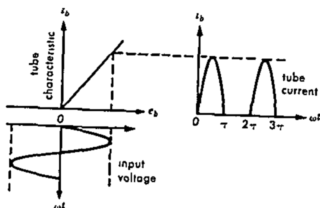


Fig 2 Rectifying action of half wave vacuum diode rectifier

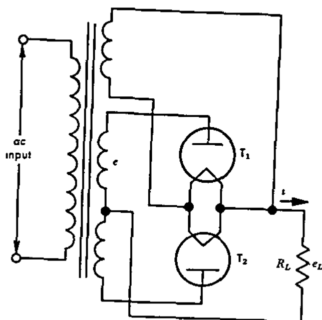


Fig 3 Full wave vacuum diode rectifier

input ac is applied to the primary of the plate transformer and the secondary voltage e supplies the rectifier tube and load resistor R_L The filament transformer supplies alternating current to heat the filament of the rectifier tube

The rectifying action of the tube is shown in Fig 2 in which the current i_b of the rectifier tube is plotted against the voltage e_b across the tube The applied sinusoidal voltage from the transformer secondary is shown under the voltage axis and the resulting current i_b flowing through the tube is shown to be half sine loops The average value of these half sine loops is the direct current flowing the ripple current is the variation of load current about the average value

Full wave rectifier circuit. The full wave rectifier circuit is shown in Fig 3 This circuit uses two separate tubes T_1 and T_2 However in most low voltage electronic supplies, both tubes are combined into one glass envelope The center tap of the transformer is connected through the load resistance R_L to the cathodes of the tubes During one-half cycle of the ac input the plate of tube T_1 is positive with respect to the cathode and T_1 conducts current from the plate to the filament This current passes through the load resistor in the direction shown by i During this time tube T_2 is not conducting since its plate is negative with respect to the cathode When the ac potential goes through zero, the plate of tube T_1 becomes negative and it stops conducting The potential on the plate of T_2 then becomes positive and T_2 starts to conduct The resulting current wave shape is shown in Fig 4

The effect of the two tubes is to produce a more continuous flow of direct current because one tube conducts for a half cycle and the second tube conducts for the second half cycle as shown in Fig 4 Comparison of Figs 2 and 4 indicates that a full wave circuit is a better rectifier than the half wave circuit

Polyphase rectifier circuits. When the dc power required by an electronic circuit is high a polyphase rectifier circuit may be used This is particularly true of power supplies for the final radio frequency and audio-frequency stages of large radio and television transmitters The rectifier tube employed in polyphase circuits generally is a gas tube that has a low voltage drop in the forward direction and thus has a high efficiency The number of phases used in these circuits is most often 3 but 2, 4 6 and 12 are used occasionally

The simplest polyphase circuit is the three-phase half wave circuit of Fig 5 The primaries of the transformers are connected in delta to the three-phase ac line, and the secondaries are connected in wye with the common connection going to one end of the load resistor The other end of the load resistor is attached to the cathodes of the three rectifier tubes required in the circuit The plates are connected to the separate ends of the three transformer secondaries

The operation of the circuit is such that tube T_1 connected to the first secondary, conducts for 120° of the ac cycle As soon as the voltage on secondary 2 equals that of secondary 1, tube T_2 starts

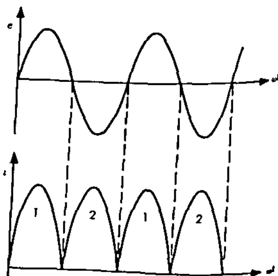


Fig 4 Applied voltage and output curre[illegible] wave rectifier

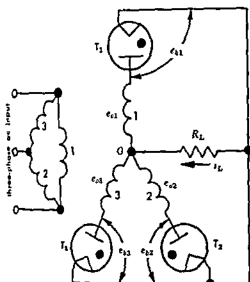


Fig 5 Three-phase half wave rectifier

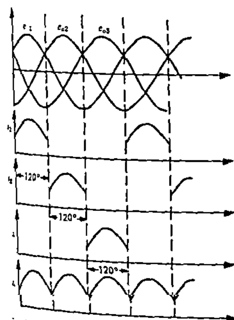


Fig. 6. Transformer voltages tube currents and load current in a three-phase half-wave rectifier

to conduct and tube T_1 stops conducting The secondary voltages e_{o1}, e_{o2} and e_{o3} are shown in Fig 6 and the tube currents are indicated as i_1 i_2 and i_3 The resulting load current i_L is also shown in Fig 6 This current is much closer to a true direct current than is the current for the single-phase circuits of Figs 1 and 3 The ripple voltage is much lower and less elaborate filter circuits are needed to smooth the output wave. The tubes used in Fig. 5 might also be ignitrons or pool cathode tank rectifiers

Another common polyphase rectifier circuit is the three phase full wave or six phase half wave circuit of Fig. 7 In this circuit the tubes conduct for 60° instead of 120° for the circuit of Fig 5 The ripple voltage for the full wave circuit is much smaller Many other polyphase rectifier circuits are possible

Bridge rectifier circuits. Bridge rectifier circuits are useful in both single-phase and polyphase applications where a transformer must be used whose secondary has no center tap or where dc voltages approximately equal to the total secondary voltage of the transformer must be obtained Another use of the bridge circuit is in ac rectifier type meters The bridge circuit is shown in Fig 8 Four separate half wave rectifier tubes are used When the left hand side of the transformer secondary is positive, current flows through tube T_1, the load resistor R_L, and tube T_3 When the secondary voltage reverses the current flows through tubes T_2 and T_4 passing through R_L in the same direction as during the first half cycle

Parallel rectifiers If greater current is desired, two or more rectifiers can be paralleled This is particularly simple if the tubes have the same ratings and are vacuum tubes For paralleling gas tubes small resistors are put in series with the tubes before they are paralleled

Controlled rectifiers Controlling the current delivered by a rectifier can be accomplished by varying the primary voltage of the power transformer or by changing a resistance in series with the load resistor The first has the disadvantage of being expensive the second leads to poor efficiency A more convenient and less expensive method is to control the angle at which the rectifier tube starts to conduct Special gas tubes that accomplish this control are thyratrons ignitrons and excitrons Thyratrons are hot cathode gas tubes with a large grid structure that prevents the arc from being ignited until the correct voltage is applied to the grid An ignitron is a cold cathode pool type tube with an igniter grid actuated by an electrical pulse The igniter of the ignitron requires a substantial amount of power which is usually supplied by an

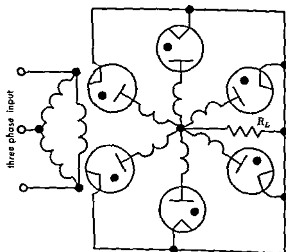


Fig 7 Three phase full wave or six phase half wave rectifier

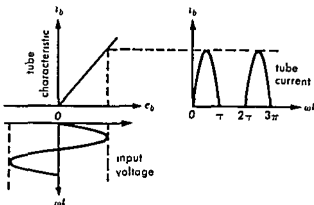


Fig 2 Rectifying action of half wave vacuum diode rectifier

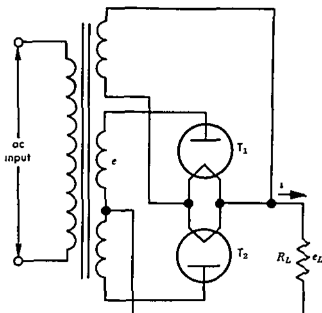


Fig 3 Full wave vacuum diode rectifier

input ac is applied to the primary of the plate transformer and the secondary voltage e supplies the rectifier tube and load resistor R_L The filament transformer supplies alternating current to heat the filament of the rectifier tube

The rectifying action of the tube is shown in Fig 2 in which the current i_b of the rectifier tube is plotted against the voltage e_b across the tube The applied sinusoidal voltage from the transformer secondary is shown under the voltage axis and the resulting current i_b flowing through the tube is shown to be half sine loops The average value of these half sine loops is the direct current flowing the ripple current is the variation of load current about the average value

Full wave rectifier circuit. The full wave rectifier circuit is shown in Fig 3 This circuit uses two separate tubes T_1 and T_2 However in most low voltage electronic supplies both tubes are combined into one glass envelope The center tap of the transformer is connected through the load resistance R_L to the cathodes of the tubes During one-half cycle of the ac input the plate of tube T_1 is positive with respect to the cathode, and T_1 conducts current from the plate to the filament This current passes through the load resistor in the direction shown by i During this time tube T_2 is no conducting since its plate is negative with respec to the cathode When the ac potential goes through zero, the plate of tube T_1 becomes negative and it stops conducting. The potential on the plate o T_2 then becomes positive and T_2 starts to conduct The resulting current wave shape is shown in Fig 4

The effect of the two tubes is to produce a more continuous flow of direct current because one tube conducts for a half cycle and the second tube conducts for the second half cycle as shown in Fig 4 Comparison of Figs 2 and 4 indicates that a full wave circuit is a better rectifier than the half wave circuit

Polyphase rectifier circuits. When the dc power required by an electronic circuit is high a polyphase rectifier circuit may be used This is particularly true of power supplies for the final radio-frequency and audio frequency stages of large radio and television transmitters The rectifier tube employed in polyphase circuits generally is a gas tube that has a low voltage drop in the forward direction and thus has a high efficiency The number of phases used in these circuits is most often 3 but 2 4 6 and 12 are used occasionally

The simplest polyphase circuit is the three-phase half wave circuit of Fig 5 The primaries of the transformers are connected in delta to the three-phase ac line and the secondaries are connected in wye with the common connection going to one end of the load resistor The other end of the load resistor is attached to the cathodes of the three rectifier tubes required in the circuit The plates are connected to the separate ends of the three transformer secondaries

The operation of the circuit is such that tube T_1 connected to the first secondary conducts for 120° of the ac cycle As soon as the voltage on secondary 2 equals that of secondary 1 tube T_2 starts

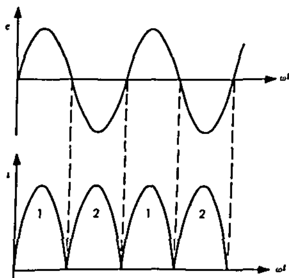


Fig 4 Applied voltage and output current of full wave rectifier

Red dwarf star

A red star of low luminosity, so designated by E Hertzsprung Dwarf stars are commonly those main sequence stars fainter than an absolute magni tude of about +1 (*see* STAR) Red dwarfs are the faintest and coldest of the dwarfs. They are present among both old and young stars Red dwarfs are

near the Sun *See* DWARF STAR [J L GR.]

Red shift

A systematic displacement towards longer wave-lengths of lines in the spectra of distant galaxies, and also of the continuous part of the spectrum First studied systematically by E Hubble red shift is central to observational cosmology, where it provides the basis for the modern picture of an ex panding universe (*see* COSMOLOGY) There are two fundamental properties of red shifts

First, the fractional red shift $\Delta\lambda/\lambda$ is independ ent of wavelength where $\Delta\lambda$ is the shift in wave-length of radiation of wavelength λ This rule has been verified from 21 cm (radio radiation from neutral hydrogen atoms) to about 6×10^{-5} cm (the visible region of the electromagnetic spec trum) and leads to the interpretation of red shift

observed effect

Second, red shift is correlated with apparent magnitude in such a way that when the former is

tax) This rule was formulated by Hubble in 1929 and the constant of proportionality bears his name. Hubble's constant is currently estimated to lie between 20 *and* 60 km/(sec) (light year) The largest red shift thus far measured corresponds to a recession speed equal to one fifth the speed of light.

The recession speed indicated by the red shift in the spectrum of a *given* galaxy is not the current value for that galaxy but the value appropriate to the epoch when the light now reaching *the Earth* was emitted Consequently the observed relation between red shift and apparent magnitude contains information about past values of Hubble's con stant, as well as about the present value If this in formation could be extracted from the *record* it would enable astronomers to choose among vari ous model universes that have been proposed by cosmologists At present however, this cannot be done. *See* EINSTEIN SHIFT [D L]

Redtop grass

tolerates both wet and dry lands and acid and infertile soils, and it is used where other species of grasses do not thrive Redtop is a perennial, spreads slowly by rootstocks, and makes a coarse loose turf Top growth is 2-3 ft tall, with moderately leafy, wiry stems *The inflorescence is a reddish* open panicle Redtop is used for pasture and hay, and is fairly nutritious if harvested promptly when heading occurs. Redtop is effective in preventing erosion by holding banks of drainage ditches, waterways, and terrace channels *See* GRASS CROPS [H B S]

Redbed

Maroon to deep-red shales and sandstones, com monly comprising thick sequences of sedimentary rocks with predominant red color, derived from sed iments deposited on land or in shallow water Red beds occur in many areas and are found throughout *the geologic column (see table)* In some areas they are eroded into spectacular badlands, as in the Painted Desert of Arizona in Bryce Canyon and Zion Canyon National Monuments in Utah, and in Roman Nose State Park in Oklahoma Rock salt, gypsum and potash salts occur in redbeds at many places and one series contains workable coal beds

The color of most redbeds is imparted by finely divided hematite (iron oxide) or by a film of hema

the erosion and deposition of such clay are apt to be redbeds Some of the coarser redbeds are red because of the presence of large quantities of the reddish mineral orthoclase feldspar Such rocks are termed arkosic The thick sequence of rich red sandstones exposed in the slopes of Sierra Diablo north of Van Horn Texas is arkosic It is one of the oldest redbeds known (Precambrian) The Gar den of the Gods in Colorado is an area of erosional forms in the arkosic Fountain sandstone, and the famous Old Red sandstone (Devonian) of England is this type of sedimentary rock *See* ARKOSE, LATERITE

Redbeds of great thickness and wide distribution have been formed at the mouths of ancient rivers in the form of deltas The Catskill delta (Devonian) of eastern New York is a famous example, similar *deposits occur in the Ordovician and Silurian of the* Appalachian Mountains and in the Permian of Ok

eral thousand feet of red clay shale and red sand stone Locally there are brown shales with abundant fresh water fish remains and in North Carolina the series contains workable coal beds The sediments were deposited by rivers carrying material into fault *troughs and their nonmarine* origin is shown by the imprint of tracks trails and raindrop impressions, and by the presence of dinosaur bones

Many redbed sequences contain evaporites—rock salt gypsum and potash salts—and clearly were

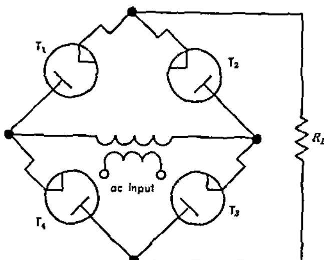


Fig 8 Single phase full wave bridge circuit

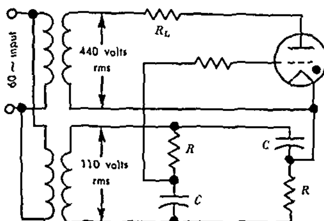


Fig 9 A phase shifting network for thyratron control

auxiliary thyratron in the control circuit The excitron resembles an *ignitron except that a* large grid structure controls the ignition of the tube *See* EXCITRON, IGNITRON, THYRATRON

One circuit for a thyratron controlled rectifier *is shown in* Fig 9 The control transformer (110-volt secondary) energizes an *RC* bridge network, which controls the phase of the grid voltage of the thyratron in relation to its plate voltage The tube will conduct when the plate is positive and when the grid voltage is increasing from negative values and is approximately equal to zero By varying either the *R* or *C* of the bridge network, the point at which the grid voltage is zero can be varied from zero degrees, when the thyratron will conduct the maximum current, to 180°, when the thyratron conducts zero current Hence, a continuous control from zero to maximum current is achieved Other control circuits also may be used

Inverse voltage. The inverse voltage of a rectifier is the voltage that the rectifier must withstand when it is not conducting or when it is conducting slightly in the reverse direction As an example, in the full wave rectifier circuit of Fig 3, when tube T_2 is conducting, tube T_1 has impressed upon it the total secondary voltage of the power transformer minus the voltage drop in T_2 For a well designed power supply the maximum value of the inverse voltage should not exceed the rated value of the rectifier specified by the manufacturer

Current ratings. Another important rating for a rectifier is the average current through it The average rectifier current of a half wave single-phase rectifier is the same as the average load current For a full wave single-phase rectifier the average rectifier current is one-half the average load current The maximum value of instantaneous current through the rectifier should not exceed the peak current rating of the rectifier This is particularly true when capacitor input filters are employed, because these filters generally produce high peak-to-average current ratios in the rectifier For information on electronic circuits in general, *see* CIRCUIT, ELECTRONIC [D L W]

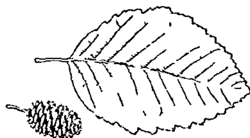
Bibliography J M Carroll, *Transistor Circuits and Applications*, 1957, H E Clifford and A H Wing, *Electronic Circuits and Tubes*, 1947, J Millman and S Seely, *Electronics*, 2d ed, 1951, S Seely, *Electronic Engineering*, 1956

Rectum

The portion of the large intestine between the anal canal and the sigmoid colon It is about 5 in in length in humans and, although "rectus" means straight, the rectum is slightly curved downward and forward from the sacral hollow to the terminal bowel The upper portion is usually empty but the lower, dilated portion forms a chamber, or ampulla, which usually contains feces and flatus. The anterior wall lies below the bladder, the posterior wall lies against the lower spine The rectum is lined with mucous membrane that is continuous throughout the intestine Control of defecation is achieved by two sphincter muscles in the terminal anal canal which is about 1½ in long *See* INTESTINE [E.G.S T]

Red alder

A deciduous tree *Alnus rubra*, which attains a *height of 60 ft* and grows from Alaska to northern California and eastern Idaho It may be recognized



Red alder, *Alnus rubra* (USDA)

by its stalked buds simple leaves and dry cone-like, ellipsoid fruit which is ½–1 in long With the bigleaf maple it shares the role of *principal* hardwood tree in the Pacific Northwest where most of the commercially important trees are conifers (*see* MAPLE) The wood is used *in furniture manufacture See* FOREST AND FORESTRY TREE [A H C]

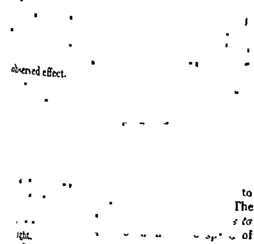
Red dwarf star

A red star of low luminosity so designated by E Hertzsprung Dwarf stars are commonly those main-sequence stars fainter than an absolute magnitude of about +1 (*see* STAR) Red dwarfs are the faintest and coldest of the dwarfs They are present among both old and young stars. Red dwarfs are the most numerous class of stars in space, although they are so faint that their presence in remote parts *of the Galaxy must be inferred from their frequency* near the Sun. *See* DWARF STAR [J L.GR]

Red shift

A systematic displacement towards longer wavelengths of lines in the spectra of distant galaxies and also of the continuous part of the spectrum First studied systematically by E Hubble, red shift is central to observational cosmology, where it provides the basis *for the modern picture of an expanding universe* (*see* COSMOLOGY) There are two fundamental properties of red shifts

First, the fractional red shift $\Delta\lambda/\lambda$ is independent of wavelength, where $\Delta\lambda$ is the shift in wavelength of radiation of wa*velength* λ *This rule has* been verified from 21 cm (radio radiation from neutral hydrogen *atoms*) *to about* 6×10^{-5} cm



The recession speed indicated by the red shift in the spectrum of a given galaxy is not the current value for that *galaxy* but the value appropriate to the epoch when the light now reaching the Earth was emitted. Consequently *the observed relation* between red shift and apparent magnitude contains information about past values of Hubble's constant, as well as about the present value If this information could be extracted from the record it would enable astro*nomers* to choose among various model universes that have been proposed by cosmologists. At present however, *this cannot* be done. *See* EINSTEIN SHIFT [D L]

Redtop grass

One of the bent grasses *Agrostis alba* and its relatives, which *occur* in cooler, *more* humid regions of the United States on a wide variety of soils Redtop tolerates both wet and dry lands and acid and infertile soils, and it is used where other species of grasses do not thrive Redtop is a perennial, spreads slowly by rootstocks, and makes a coarse loose turf Top growth is 2–3 ft tall, with moderately leafy, wiry stems The inflorescence is a reddish open panicle Redtop is used *for pasture and hay*, and is fairly nutritious if harvested promptly when heading *occurs Redtop is effective in preventing* erosion by holding banks of drainage ditches, waterways, and terrace channels *See* GRASS CROPS [H B S]

Redbed

Maroon to deep-red shales and sandstones, commonly comprising thick sequences of sedimentary rocks with predominant red color, derived from sediments *deposited on* land *or in* shallow water Red beds occur in many areas and are found throughout *the geologic* column (see table) In some areas they are eroded into spectacular badlands, as in the Painted Desert of Arizona, in Bryce Canyon and Zion Canyon National Monuments in Utah, and in Roman Nose State Park in Oklahoma Rock salt, *gypsum and pota*sh *salts occur in* redbeds at many places and one series contains workable coal beds

The color of most redbeds is imparted by finely divided hematite (iron oxide) or by a film of hema

the erosion and deposition of such clay are apt to be redbeds Some of the coarser redbeds are red because of the presence of large quantities of the reddish mineral orthoclase feldspar Such rocks *are termed arkosic The thick sequence of rich red* sandstones exposed in the slopes of Sierra Diablo north of Van Horn Texas, is arkosic It is one of the oldest redbeds known (Precambrian) The Garden of the Gods in Colorado is an area of erosional forms in the arkosic Fountain sandstone, and the famous Old Red sandstone (Devonian) of England *is this type of sedimentary rock See* ARKOSE, LATERITE

Redbeds of great thickness and wide distribution have been formed at the mouths of ancient rivers in the form of deltas The Catskill delta (Devonian) of eastern New York is a famous example, similar deposits occur in the Ordovician and Silurian of the *Appalachian Mountains* and in the Permian of Oklahoma The Newark series of the Connecticut River Valley northern New Jersey *eastern Pennsylvania*, central Virginia and North Carolina consists of several thousand feet of red clay shale and red sandstone Locally there are brown shales with abundant fresh water fish remains, and in North Carolina the series contains workable coal beds The sediments were deposited by rivers carrying material into fault troughs and their nonmarine origin *is shown by the* imprint of tracks trails and raindrop impressions, and by the presence *of dinosaur bones*

Many redbed sequences contain evaporites—rock salt gypsum and potash salts—and clearly were

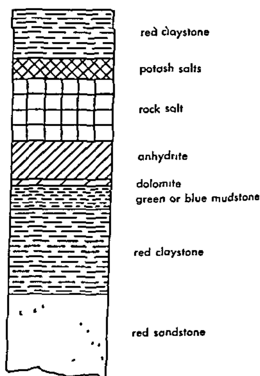
Geographic and geologic occurrence of redbeds

Geologic column	Redbed formation or series	Geographic location	Characteristics of deposition and deposits
Cenozoic			
Quaternary	Laterite soils	Humid tropics	Iron rich red clay, formed by residt weathering
Tertiary	Wasatch formation	Utah Wyoming	Nonmarine gravel, sand clay wı lenses of lake deposits
Mesozoic			
Cretaceous	Djadokhta sandstone	Mongolia	Red eolian sand containing dinosaı remains
Jurassic	Wingate Navajo	Colorado Plateau	Thick intensely cross bedded red san stones
Triassic	Newark series, Chugwater group	Eastern U S Wyoming	Red shale and sandstone, gypsur Newark deposited in fault trougl
Paleozoic			
Permian	Hennessey shale	Central Oklahoma	Red shale of deltaic origin
	Salado formation	SE New Mexico	Contains rock salt and potash sal
	Blaine formation	Oklahoma Texas, Kansas	Red shale, gypsum beds marine origı
	Wellington formation	Kansas and Oklahoma	Red shale rock salt
Pennsylvanian	Fountain sandstone	Colorado	Arkosic sandstone and conglomeral
Mississippian	Mauch Chunk formation	Pennsylvania	Deltaic varicolored shale
Devonian	Catskill formation	Eastern New York	Thick, coarse grained, deltaic
Silurian	Bloomsburg formation	Pennsylvania	Red sandstones of deltaic origin
Ordovician	Queenston shale	Niagara Gorge	Deltaic red shale
	Juniata shale	Appalachian Mts	Deltaic red shale
Cambrian	Flathead sandstone	Wyoming	Iron stained quartzitic sandstone
Precambrian	Hazel sandstone	West Texas	Arkosic iron bearing sandstone

deposited in saline basins or restricted arms of the sea Rock salt is mined from Permian redbeds in Kansas, gypsum is mined in Oklahoma, and potash salts are produced from mines near Carlsbad, New Mexico Redbeds of this type normally were deposited in a cyclical pattern starting with sandstones and shales and terminating in potash salts (see illustration) The formation from which rock salt is mined in Kansas (Wellington formation) is predominantly red clay shale, but it contains gray and greenish gray shales, tongues of reddish sandstone, and conglomerates containing bone fragments At places it includes salt lake deposits of varicolored clay shales as well as dolomite lenses containing fossil insects, conchostracans and xiphosurans *See* EVAPORITE (SALINE), GYPSUM, ROCK SALT

A bright red sandstone in Mongolia is composed of cemented wind blown sand derived from an older red sandstone The formation contains fine specimens of primitive horned dinosaurs and dinosaur eggs now on display in the American Museum of Natural History in New York Perhaps too, of eolian origin are the thick red sandstones that form the spectacular cliffs in Zion National Monument and adjacent areas The pigment in these redbed sandstones is a thin coating of iron oxide on each grain, a coating that must have been introduced after deposition

Other types of red rocks have been formed under special conditions but are not classified as red beds under the present definition Thick sedimentary iron ores such as the Clinton hematite of the Appalachian region and the Precambrian hematite ore of the Mesabi district in Minnesota are largely composed of red iron oxide The red color of a limestone in the Grand Canyon is a stain extending only a few feet below the surface The reddish Cambrian sandstones of central United States and the ancient



Sequence of redbeds with evaporites characteristic of Permian formations in southwestern United States

red rocks of Northern Michigan are iron stained Some red clays, one of which occurs in Timor, are believed to have been deposited in sea deeps [C C BR]

Redstart

Either of two small American wood warblers, both of striking coloration The breeding male of the American redstart *Setophaga ruticilla*, has glossy





The redstart *Setophaga ruticilla*, length to 5¾ in (From E L Palmer *Fieldbook of Natural History* McGraw Hill 1949)

black upper parts, throat and breast, with orange patches on the shoulder, across the wing, and on the basal two-thirds of the outer tail feathers, the belly is white The female is grayish olive with yellow replacing the orange

The painted redstart S picta glossy bl [illegible]

[illegible] of the pine oak woods of the Southwest, whereas *S ruticilla* is found in open woodlands throughout much of the United States and Canada *See* PASSERIFORMES [J D B]

Redwood

A member of the pine family, *Sequoia sempervirens* the tallest tree in America It attains a height of 350 ft and a diameter of 27 ft Its present range is limited to a strip along the Pacific Coast about



Redwood Sequoia sempervirens (USDA)

35 miles wide and 500 miles long extending from southwest Oregon to about 100 miles south of San Francisco The leaves are evergreen, sharply pointed, small (⅓–1 in long), distichous (disposed in two vertical rows) on short branches and scalelike on the main stem The cones are egg shaped, about 1 in long and ½ in broad The bark is a dull red brown, on old trees sometimes 1 ft thick, densely fibrous and highly resistant to fire The tree gets its common name from the color of the bark as well as that of the heartwood The wood is similar in weight to that of eastern white pine and is likewise strong and stiff, easy to work, warps or swells but little, and resists attacks of fungi and insects (*see* PINE) It is used for bridge timbers, tanks flumes, silos, posts, shingles, paneling, doors, caskets, furniture, siding and many other building purposes The present stand of saw timber is estimated at 40,000 000 000 board ft The annual cut is about 800 000,000 board ft *See* FOREST AND FORESTRY, TREE [A H C]

Reef

A mass or ridge of rock or rock forming organisms in a water body Usually the term reef means a rocky menace to navigation within 6 fathoms of the



Seaward reef and surge channels, Bikini Atoll (*Official photograph USGS*)

water surface Various kinds of calcium carbonate–secreting animals and plants create organic reefs throughout the warmer seas Naturally cemented sand ridges make reefs along the coast of Brazil and elsewhere Rocky shores of seas lakes and navigable rivers commonly exhibit reefs of rock types similar to those of the adjacent land *See* ATOLL BARRIER REEF, CORAL REEF, FRINGING REEF ORGANIC REEF [P E C]

Reentry

The return of a ballistic flight vehicle into the earth s atmosphere Preliminary reentry experience has been with nose cones as they descend The nose cone is the part of a ballistic missile which contains the payload Motion and heating of the nose cone are of principal concern during entry, and these phenomena are governed primarily by the velocity and angle of entry into the atmosphere the properties of the atmosphere and the shape and weight of the nose cone

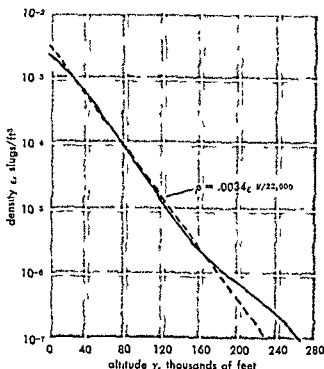


Fig 1 Angle and velocity of entry

Atmosphere entry. The velocity and angle of entry are related to the range of flight in the manner shown in Fig 1 To minimize the size of boosters required to propel a nose cone over a certain range it is desirable to keep the initial projection velocity as low as possible *Figure 1 shows that in this event* the angle of entry into the earth's atmosphere de creases continuously with increasing range from a value near 45° for very short range to a value near 0° for very long range The maximum velocity re quired under these circumstances increases to the order of satellite speed or 18 000 mph For the payload to return accurately to a desired location or target, it is desirable that it maintain large entry angles with the horizontal, consequently velocities significantly in excess of satellite speed *may be required for very long range flight*

The properties of the earth's atmosphere which bear most importantly on motion and heating during entry are composition, density, *and distribution of* density in the atmosphere In the lower portions of the atmosphere where motion and heating of a nose cone can be significant, the composition is essen tially the same as that at sea level, about 20% molecular oxygen and 80% molecular nitrogen

The thermal and transport properties of air of this composition, including such quantities as the specific heats, viscosity, and thermal conductivity are well known at normal temperatures, but the temperatures in the air disturbed by the nose cone *in its high speed* flight through the atmosphere may reach many thousands of degrees Fahrenheit In this event molecular oxygen and to some extent molecular *nitrogen tend to be* dissociated into atoms and some ionization occurs The thermal and trans port properties of dissociated air are not so well established as those for *undissociated air*, although substantial progress is being made on this problem with the aid of theoretical quantum mechanical studies and with *the aid of experiments* in ‹ tubes The density of air in the earth's lower at phere varies with altitude in the manner show Fig 2 Density decreases rapidly and more or exponentially with increasing altitude Accord in considerations of motion and heating of cones during entry, the assumption of an expo tial atmosphere is frequently employed

Deceleration. The key *factor in the motion* nose cone during atmosphere entry is the dece tion due to the action of aerodynamic drag M mum *deceleration* varies with range in the ma shown in Fig 3 This deceleration first incre with increasing range, reaching a peak of al 60 g at a range of about 5000 miles, and th decreases to the order of 10 g for ranges of a 12,500 miles and greater The increasing dece tion up to a range of 5000 miles is caused dominantly by the increasing entry velocity *decreasing deceleration with range over 5000 r* is caused predominantly by the decreasing e angle which permits the missile to slow down i gradually in the atmosphere Higher decelerat can be expected at longer ranges if entry angle increased to achieve higher target accuracy In case it is clear that maximum deceleration is dependent on shape or, for that matter, on si weight of the nose cone Changing these quant will, to be sure change the altitude in the at phere where maximum deceleration occurs, bu *nose cones of practical interest, such changes* not alter the magnitude of the deceleration high deceleration in turn implies that the struc of the nose cone and payload must be designe

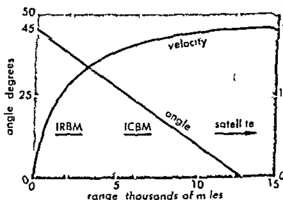


Fig 2 Variation of density with altitude

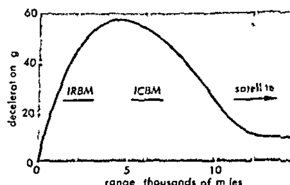


Fig 3 Maximum deceleration during entry

withstand large aerodynamic drag forces A far more severe design requirement, however, is im posed by the severe heating experienced by a nose cone during entry See MISSILE, NOSE CONE, SPACECRAFT STRUCTURE [A J EG]

Reflection (electromagnetic radiation)

The return of electromagnetic waves from surfaces on which they are incident When radiation crosses a boundary between two media and suffers a change in velocity a fraction R of its energy is reflected The magnitude of R depends on the difference in velocity and the angle of incidence

tion is different for each material, depending on the atomic molecular and crystal structure The ratio of velocity in vacuum to that in a given mate rial is called the refractive index (index of refrac tion) of the material It is found to vary with the wavelength of the radiation and can be represented by real (noncomplex) numbers for those wave length regions sufficiently removed from absorp tion bands in the material See REFRACTION OF WAVES.

material a *complex* refractive index $\mu - ik$ the real

take up energy from the electromagnetic field For metals, in which the electrons are free to increase

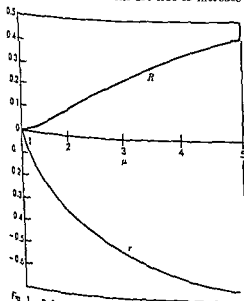


Fig 1 Reflected amplitude r and energy R versus refractive index μ from a r-d electric interface for normal incidence.

their energy, radiation of all wavelengths can be absorbed The rate at which energy is absorbed determines the reflectivity, and is determined by the number of electrons available in the metal See ABSORPTION (ELECTROMAGNETIC RADIATION)

The present article is concerned with specular reflection only, that is, reflection from surfaces which are optically smooth, and neglects effects of diffuse reflection and scattering

REFLECTION FROM DIELECTRICS

Normal incidence reflectivity. When radiation is incident along a normal to an interface between two nonabsorbing media the ratio of the reflected to the incident energy is

$$r = \left(\frac{\mu - \mu'}{\mu + \mu'}\right) \qquad (1)$$
$$R = r^2$$

where R = fraction of energy reflected

This expression the standard Fresnel reflection law indicates the phase shift on reflection to be zero when the incident medium has the higher in dex, and 180° (since r is negative) otherwise

For many applications the incident medium is air with $\mu \cong 1$ giving

$$r \cong \frac{1 - \mu'}{1 + \mu'} \qquad (2)$$

The curve $(1 - \mu')/(1 + \mu')$ vs μ' is part of a branch of a hyperbola for positive values of μ' For many common optical materials, including most glasses $\mu' \cong 1.5$ giving $r = -0.20$ and $R = 0.04$ Therefore approximately 4% of the incident light is reflected from a single glass surface See OPTICAL MATERIALS

Many of the materials transparent in the in frared spectral regions have much higher refractive indices Germanium transparent for wavelengths greater than 1.8 microns has $\mu' = 4$ $r = -0.6$, $R = 0.36$ This high reflectivity seriously reduces its transmission unless the surfaces are coated with nonreflecting films

The variation of reflectivity with refractive index for normal incidence in air is given in Fig 1

Plane parallel interfaces If radiation is incident on a stratified medium (one which contains parallel interfaces for example a pile of plates), a simple relation holds for the reflectivity of the medium in terms of the reflectivity of the individual inter faces If r and t_i are the fractions of reflected and transmitted energy for the p individual interfaces and if R and T are the corresponding values for the entire medium then

$$\frac{R}{T} = \sum_{i=1}^{p} \frac{r_i}{t_i} \qquad (3)$$

Here the r_i are computed from Eq (1) This rela tionship includes the multiple reflections that occur in the medium but cannot be used if interference

effects redistribute the energy between the reflected and transmitted waves

Thin films Many interesting reflection effects can be produced by taking advantage of interference between the multiply reflected components of a beam of light in thin films The reflectivity from a dielectric may be made to take on any value from zero to nearly 100% through the use of one or more thin, nonabsorbing layers deposited on its surface

To reduce the reflectivity to zero, a single thin film of index μ_1 may be deposited on a dielectric of index μ_2 to a thickness d_1 such that

$$\mu_1 d_1 = (2n - 1)\frac{\lambda}{4} \qquad \mu_1 = (\mu_0\mu_2)^{1/2} \tag{4}$$

where n = an integer
λ = wavelength of the radiation
μ_0 = refractive index of incident medium (if air $\mu_0 \cong 1$)

Under these conditions the reflectivity is the same at each of the two interfaces The multiply reflected beams r_2 r_3, r_4, , of Fig 2 are all in phase each one traveling an additional ½ wave through twice the film thickness more than the previous beam and in addition suffering a phase shift equivalent to a ½-wave displacement at the second interface The resultant amplitude of these beams is, therefore,

$$\begin{aligned} r_2 + r_3 + r_4 + \quad &= t^2 r(1 + r^2 + r^4 + \quad) \\ &= \frac{t^2 r}{1 - r^2} \\ &= r \end{aligned} \tag{5}$$

since $r^2 + t^2 = 1$, that is the transmitted and reflected beams contain all the energy The first beam r_1 has an amplitude $-r$ exactly 180° out of phase with the other beams because of its phase shift on reflection Therefore the resultant reflected amplitude is zero no energy is reflected and 100% transmission occurs

The general equation for the amplitude of a wave reflected from a single quarter wave film on a substrate is

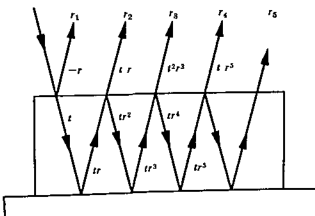


Fig 2 Multiply reflected beams from a thin film (shown separated for clarity)

Reflected amplitude of quarter wave film (index μ_1) on glass (index 1 5)

μ_1	$t_1/t_2 = \mu_1^2/\mu_0\mu_2$	$r = \frac{1-(t_1/t_2)}{1+(t_1/t_2)}$
1 0	0 67	0 2
1 2	0 95	0 03
1 23	1 0	0 0
1 4	1 3	−0 13
1 6	1 7	−0 26
1 8	2 2	−0 38
2 0	2 7	−0 46
2 5	4 2	−0 62
3 0	6 0	−0 71
4 0	10 7	−0 83
5 0	16 7	−0 89
6 0	24	−0 92

$$r = \frac{1 - (t_1/t_2)}{1 + (t_1/t_2)} \quad \text{or} \quad \frac{1-r}{1+r} = \frac{t_1}{t_2} \tag{6}$$

where $t_1 = \mu_1/\mu_0$ and $t_2 = \mu_2/\mu_1$ The accompanying table gives t_1/t_2 as a function of μ_1 and r as function of t_1/t_2 Knowing μ_2 and μ_0, t_1/t_2 and may be determined as a function of μ_1

The use of a ¼ wave film whose index is greater than that of the substrate results in an increase reflectivity according to Eq (6) By this means, glass surface reflecting 4% can be made to reflect 32% by depositing ¼ wave of ZnS, index 2 3

If wavelengths λ, different from that for which the film has an optical thickness of ¼ wave are incident, the reflected amplitude will vary according to the more general relationship

$$\begin{gathered} \frac{1-r}{1+r} = \frac{t_1(t_2 - i\tan\delta_1)}{1 - it_2\tan\delta_1} \\ \delta_1 = \frac{2\pi}{\lambda}\mu_1 d_1 = (n - \tfrac{1}{2})\pi\frac{\lambda_0}{\lambda} \end{gathered} \tag{7}$$

Solving this equation for the complex value of r one can determine the phase shift on reflection and also the reflected energy

For the special case of a film ½ wave thick that is

$$\begin{gathered} \mu_1 d_1 = n\frac{\lambda_0}{2} \\ \delta_1 = n\pi \quad \text{and} \quad \frac{1-r}{1+r} = t_1 t_2 = \frac{\mu_2}{\mu_0} \end{gathered} \tag{8}$$

This is identical to the reflectivity obtained without the ½-wave layer present Therefore a ½-wave layer can be disregarded in reflection computations

The reflectivity of a dielectric surface can be increased to as close to 100% as one wishes by depositing alternating ¼ waves of high and low index materials The peak reflectivity can be calculated by assigning to the mth interface in the pile (counting down from the air interface as 1) a value

$$t_m = \frac{\mu_m}{\mu_{m-1}} \tag{9}$$

where μ_m is the refractive index of the mth layer

and μ_{m-1} that of the layer above If there are n layers there are $n+1$ interfaces (Fig 3)

$$\frac{1-r}{1+r} = \frac{t_1 t_3 t_5 \ldots}{t_2 t_4 t_6 \ldots}$$

$$= \frac{\mu_1^2 \mu_3^2 \mu_5^2 \ldots}{\mu_2^2 \mu_4^2 \mu_6^2 \ldots} \mu_s \quad \text{for an even number of layers} \qquad (10)$$

$$= \frac{\mu_1^2 \mu_3^2 \mu_5^2 \mu_7^2 \ldots}{\mu_2^2 \mu_4^2 \mu_6^2 \ldots \mu_s} \quad \text{for an odd number of layers}$$

If the alternate layers are identical $\mu_1 = \mu_3 = \mu_5 \ldots$ and $\mu_2 = \mu_4 = \mu_6 = \ldots$

$$\frac{1-r}{1+r} = \frac{\mu_1^n}{\mu_2^n}\mu_s \quad n \text{ even}$$
$$= \frac{\mu_1^{n+1}}{\mu_2^{n-1}\mu_s} \quad n \text{ odd} \qquad (11)$$

For wavelengths both greater and smaller than λ_0 (the wavelength for which the layers are ¼ wave) the reflectivity falls off only slightly until a critical wavelength is reached beyond which the drop is more rapid The critical wavelengths λ are dependent on *the ratio of refractive indices of the* two materials, their values being given by

$$\cos^2\left(\frac{\pi \lambda_0}{2 \lambda}\right) \quad \left[\frac{1 - (\mu_2/\mu_1)}{1 + (\mu_2/\mu_1)}\right]^2 \qquad (12)$$

The amount by which the *reflectivity increases* as an additional pair of layers is deposited increases according to the departure of μ_1/μ_2 from unity As the *number of layers is increased the* curve of reflectivity vs wavelength near λ_0 takes on the character of a square wave whose width decreases as $\mu_1/\mu_2 \rightarrow 1$ In the vicinity of the wavelength for which the layers are ¾ waves that is around $\lambda_0/2$ the curve has an undulating form

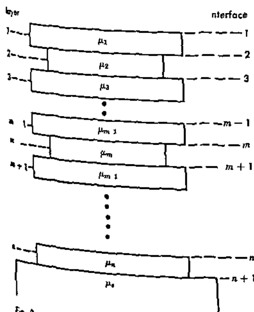


Fig 3 Arrangement for multilayers

taking on a value equal to the reflectivity of the bare substrate at $\lambda_0/2$

The particular advantages of utilizing *multilayer* films for filters lie in their efficiency (the energy not reflected is transmitted) the steepness of the slope $(dR/d\lambda)$ and the possibility of depositing films to operate at almost any wavelength desired *See* INTERFERENCE FILTER OPTICAL

Oblique incidence In the case of nonnormal incidence for dielectric materials the direction of propagation of the energy *is no longer the same* for all materials but varies according to Snell s law At each interface between media of different refractive indices μ_1 and μ the light is deviated the angles θ_1 and θ_2 made with the normal to the interface being related by

$$\mu_1 \sin \theta_1 = \mu_2 \sin \theta_2 \qquad (13)$$

If light is transmitted through a set of plane parallel interfaces the direction of propagation in any medium is the same as it would be if the layers above it were removed *It depends only on the* angle of incidence in the incident medium and the ratio of its refractive index to that of the incident medium

Polarized components Electromagnetic radiation also displays its characteristics as a transverse wave by separating into two polarized components when incident at oblique angles They are conventionally considered to separate according to the plane in which the electric wave is located The plane of incidence is that containing the direction of propagation and the normal to the surface at the point of contact That part of the radiation with its resultant electric wave in this plane is designated as p polarized (parallel) and the part vibrating perpendicular to this plane is called s polarized

Each component has its own reflection characteristics Fresnel s equations for the two components are

$$\frac{1 - r_s}{1 + r_s} = \frac{\mu_2 \cos \theta_2}{\mu_1 \cos \theta_1} = t_s \quad (s \text{ polarization})$$
$$\frac{1 - r_p}{1 + r_p} = \frac{\mu_1 \cos \theta_2}{\mu_2 \cos \theta_1} = t_p \quad (p \text{ polarization}) \qquad (14)$$

These equations predict the reflectivity in particular under three interesting conditions At grazing incidence $\cos \theta_1$ vanishes and the right side of *both expressions* becomes infinite indicating 100% reflection for each polarization

When θ_1 and θ_2 are complementary the expression for the p polarization becomes unity making r_p vanish. The angle of incidence $\theta_B = \tan^{-1}(\mu_2/\mu_1)$ is called the polarizing angle or Brewster s angle The reflected and refracted waves are propagated at right angles to each other the reflected wave containing only s polarization *See* POLARIZED LIGHT

The third interesting condition occurs when $\theta_2 = \pi/2$ and $\mu_1 > \mu_2$ In this case, *both expressions* vanish because $\cos \theta_2 = 0$ and both components are completely reflected Since this occurs

only on reflection from a high to a low index medium, it is called internal reflection The angle of incidence is designated as the critical angle θ_c and satisfies the relation $\sin \theta_c = \mu_2/\mu_1$ For angles of incidence greater than this, $\cos \theta_2$ becomes an imaginary quantity Solving for r_s and r_p yields complex expressions whose amplitude is unity (indicating 100% reflectivity), but whose arguments range from 0 to π, indicating the phase shifts on reflection experienced by polarization

Although no energy is transmitted into the lower index medium, an attenuated wave is progagated along the interface, its amplitude falling off exponentially with d, the distance from the interface The presence of this wave can be detected when the low index medium is air Total reflection can be destroyed by bringing an object close to, but not touching the interface, thereby disturbing the attenuated wave

The reflectivity for the two polarizations is shown in Fig 4 for incidence from a low to high index and high to low index interface

Thin films Reflectivity from thin films for oblique incidence can be treated similarly to the case for normal incidence, where one uses the t's for each polarization at each interface, as defined in Eqs (14) A further difference occurs in the effective optical thickness of a layer at oblique incidence According to the interference produced by a thin film, its effective thickness is reduced by $\cos \theta$, where θ is the propagation angle in the medium, determined from Eq (13) For example, a ½-wave film at normal incidence would have an effective optical thickness of only ¼ wave when light traversed it at a 60° angle

Because of this effect, a multilayer filter in which the layers are matched at one angle of incidence will become mismatched at another Each layer will change effective thickness according to its index which, through Snell's law, determines the angle In general, the effect of this mismatching will be to reduce the peaks and slopes of the reflectivity vs wavelength curve

Selective reflection from crystals. The discussion to this point has been concerned with reflectivity of nonabsorbing media far removed from absorption bands These bands are located in the spectral regions where the frequency of the radiation corresponds to a resonance frequency of the atoms, molecules, or crystal lattice of the medium Since this radiation is strongly absorbed it is also strongly reflected The metallic sheen of dye crystals, which have very strong absorption bands in the visible spectrum, is caused by selective reflection Crystalline solids such as rock salt or quartz the lattices of which are built up of atoms bearing net electric charges, show strong selective reflection in the infrared region at wavelengths near those of the strong absorption bands associated with lattice vibrations in the crystal By reflecting an infrared beam several times from such a material, highly monochromatic radiation can be obtained at the specific wavelengths These monochromatic beams are referred to as residual rays or reststrahlen Figure 5 indicates the residual rays for some common crystals *See* IONIC CRYSTALS

REFLECTION FROM METALS

In determining the reflective properties of metallic substances, the solutions derived for dielectrics apply, but the refractive indices and the angles determined from the indices through Snell's law become complex The reflectivity of some common metals is given in Fig 6 The reflectivity falls off

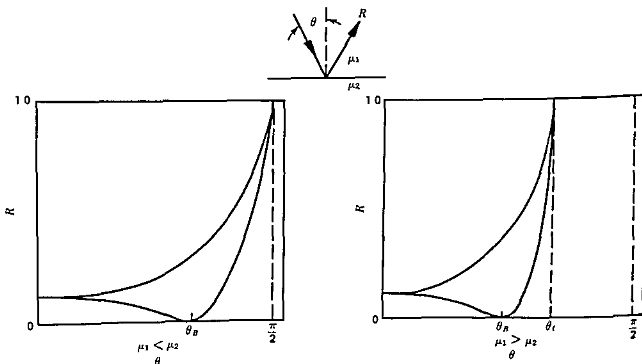


Fig 4 Reflectivity vs angle of incidence for dielectrics

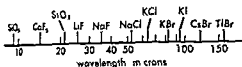


Fig 5. Wavelengths of residual rays for various crystals.

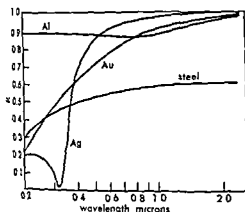


Fig 6 Metallic reflection

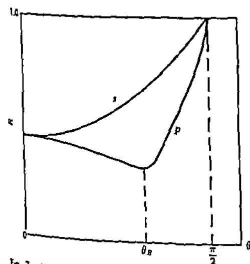


Fig 7 Metallic reflectivity vs angle of incidence

toward short wavelengths a characteristic of most metals, indicating a reduction in the number of electrons capable of absorbing this energy Of particular interest is the "window" for silver at 3200 A The transmission through this region becomes high enough to allow thin films to be used for isolating a spectral region approximately 100 A wide.

Typical curves of the reflectivity of the polarized components versus angle of incidence are given in Fig. 7 showing the similarity with the curves for dielectrics (Fig. 4) The values of the optical constants μ and k can be derived from experimental data determining the angle of incidence for the minimum reflectivity of the p polarization that is, the polarizing angle and the ratio of r_p/r_s at this angle. The experimental procedure is facilitated by the fact that the difference in phase shift on reflection for the two components is exactly $\pi/2$ at this angle Thus, the polarizing angle can be accurately determined by using a ¼ wave plate to convert the elliptically polarized light to linear polarization For any other angle of incidence, linear polarization will not result, making it impossible to produce extinction for any orientation of an analyzer

An interesting property of metallic reflection occurs when an extremely thin film (less than 100 A thick) is deposited on a transparent substrate Although the reflectivity from the air side may be as high as 20%, that on the glass side is effectively reduced to zero Contrary to the dielectric nonreflecting film, the low reflectivity extends over a large part of the visible spectrum With a film this thin it is not the optical thickness that plays the predominant part in the interference effects, but the phase shift on reflection at the air metal and metal substrate interfaces If the back surface reflection of a beam splitter in a noncollimated beam gives a doubling of an image, it can be more completely eliminated through the use of a thin metal film than a ¼ wave dielectric film *See* ALBEDO, MIRROR OPTICS, OPTICS, GEOMETRICAL, REFLECTION (SOUND), REFLECTION AND TRANSMISSION COEFFICIENTS [H D PO]

Bibliography F A Jenkins and H E White, *Fundamentals of Optics*, 3d ed, 1957, J A Stratton *Electromagnetic Theory*, 1941, J Strong, *Concepts of Classical Optics*, 1958

Reflection (sound)

The return of sound waves from surfaces on which they are incident Suppose a sound wave strikes a smooth surface that is large compared to the wavelength of the sound and suppose that the path of this sound wave is represented by a ray that is by a line perpendicular to the advancing wavefront By the law of reflection the angle of reflection r for this ray equals the angle of incidence i, and the reflected ray lies in the plane of incidence Figure 1 shows reflection from a plane surface

The geometrical laws for reflection of sound waves are the same as those for light waves The

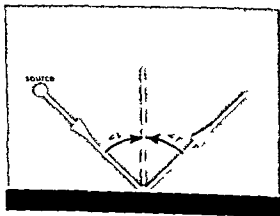


Fig 1 Reflection from a plane surface

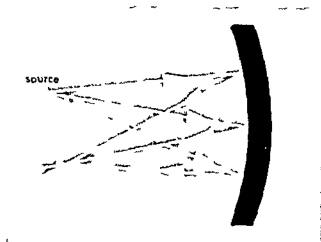


Fig 2 Reflection from a concave spherical surface

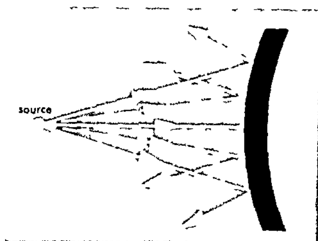


Fig 3 Reflection from a convex surface

apparent differences involve only questions of scale, because the average wavelength of sound is about 100 000 times that of light For example a mirror or lens used to produce a beam of sound waves must be enormously large compared to mirrors and lenses used in optical systems *See* REFLECTION (ELECTROMAGNETIC RADIATION)

A concave surface tends to concentrate the reflected sound waves as illustrated in Fig 2 Such surfaces are sometimes used to advantage as reflectors but if used indiscriminately they may lead to poor acoustics as a result of undesirable focusing effects Convex reflectors (Fig 3) tend to spread the reflected waves Therefore, when placed at the boundaries of a room, they tend to diffuse the sound throughout the room For this reason some radio broadcasting studios employ cylindrical convex panels as part of their wall construction to promote diffusion

The law of reflection is often used in investigating the effects of various shapes of a proposed room on the distribution of sound in that room Such studies can lead to the design of interior surfaces that will give beneficial reflections, or to the elimination or modification of surfaces that otherwise would give rise to echoes However, caution must be exercised in such applications of the law of reflection, because the wave properties of sound are neglected in this simplification of the behavior of sound For a discussion of sound reflections in water, *see* UNDERWATER SOUND *See also* ARCHITECTURAL ACOUSTICS, ECHO, SOUND [C M H]

Reflection and transmission coefficients

When an electromagnetic wave passes from a medium of permeability μ_1 and dielectric constant ϵ_1 to one with values μ_2 and ϵ_2, part of the wave is reflected at the boundary and part transmitted The ratios of the amplitudes in the reflected wave and the transmitted wave to that in the incident wave are called the reflection and transmission coefficients, respectively For oblique incidence, the reflection and refraction formulas of optics are most convenient, but for normal incidence of plane waves on plane boundaries, such as occur with transmission lines, wave guides, and some free waves, the concept of wave impedance and characteristic impedance is useful

For a z directed wave with electric intensity **E** in the x direction and magnetic intensity **H** in the y direction, the total phasor fields on the incident side are

$$\check{E}_x = E_0 e^{-jkz} + \check{E}'_0 e^{jkz} \qquad (1)$$

$$\check{H}_y = (\eta)^{-1}(E_0 e^{-jkz} - E'_0 e^{jkz}) \qquad (2)$$

where primes are used for reflected quantities, and η is the wave impedance The sign difference in Eqs (1) and (2) is due to the fact that Poynting's vector $\frac{1}{2}\mathbf{E} \times \mathbf{H}$ is positive for the incident and negative for the reflected wave (*see* ELECTROMAGNETIC RADIATION, POYNTING S VECTOR) For the transmitted wave

$$\check{E}''_x = E''_0 e^{-jkz} \qquad \check{H}'_y = (\eta'')^{-1} E''_0 e^{-jkz} \qquad (3)$$

Since the tangential components of **E** and **H** are continuous across the boundary at $z = 0$, $\check{E}_x = \check{E}''_x$ and $\check{H}_y = \check{H}''_y$, so that

$$E_0 + E'_0 = E''_0 \qquad \eta''(E_0 - E'_0) = \eta E'_0 \qquad (4)$$

The ratios for the reflected and transmitted fields obtained by solving these equations are

$$\frac{E'_0}{E_0} = \frac{\eta'' - \eta}{\eta'' + \eta} \qquad \frac{E''_0}{E_0} = \frac{2\eta''}{\eta'' + \eta} \qquad (5)$$

These are the reflection and transmission coefficients respectively

Coefficients for optics. Equation (5) holds for normal incidence in optics, if the velocities v and v'' are written for η and η'' For a plane wave whose electric vector is normal to the plane of incidence and whose direction makes an acute angle θ with the normal to the interface the reflection and transmission coefficients are

$$\frac{E'_0}{E_0} = -\frac{\sin(\theta - \theta'')}{\sin(\theta + \theta'')} \qquad \frac{E''_0}{E_0} = \frac{2\sin\theta''\cos\theta}{\sin(\theta + \theta'')} \qquad (6)$$

where $v'' \sin\theta = v \sin\theta''$ When the electric vector lies in the plane of incidence, the coefficients become

$$\frac{E'_0}{E_0} = \frac{\tan(\theta - \theta'')}{\tan(\theta + \theta'')} \qquad \frac{E''_0}{E_0} = \frac{2 \sin\theta'' \cos\theta}{\sin(\theta + \theta'')\sin(\theta - \theta'')} \tag{7}$$

The ratio v/v'' is the index of refraction See REFRACTION OF WAVES

Wave guides In wave guides, as in free space, the characteristic impedance is defined as the ratio of the transverse electric field E_t to the transverse magnetic field H_t For wave guides, this ratio depends on the frequency and the dimensions of the wave guide as well as on the permeabilities and dielectric constants For a transverse interface, the boundary conditions used for Eq (4) on the tangential fields still hold Thus, Eq (5) for the reflection and transmission coefficients is valid if η and η'' are replaced by the characteristic impedances on the incident and emergent sides respectively

Transmission lines and networks Let $\check Z$ and $\check Z''$ be the characteristic impedances on the incident and emergent sides of a discontinuity in a transmission line or on the two sides of a junction between two networks Then the relations between the potentials and currents of the incident, reflected, and transmitted waves are respectively

$$\check V = \check Z \check I \qquad \check V' = -\check Z \check I' \qquad \check V'' = \check Z'' \check I'' \tag{6}$$

At the discontinuity potential and current must be continuous so that

$$\check V + \check V' = \check V'' \qquad \check I + \check I' = \check I'' \tag{7}$$

Solution for the ratios gives

$$\frac{\check V'}{\check V} = \frac{\check Z'' - \check Z}{\check Z'' + \check Z} \qquad \frac{\check V''}{\check V} = \frac{2\check Z''}{\check Z'' + \check Z} \tag{8}$$

These are the reflection and transmission coefficients See TRANSMISSION LINES

Coefficients for acoustics. *Equation (8) holds* in acoustics provided acoustic impedance is substituted for electrical impedance Acoustic impedance is defined as the product of the density of a medium by the speed of sound in it There are two types of waves in solid mediums longitudinal waves and shear waves and thus there are two impedances. See IMPEDANCE ACOUSTIC [W R SM]

Bibliography D E Gray (ed) *American Institute of Physics Handbook* 1957 S Ramo and J R Whinnery *Fields and Waves in Modern Radio* 2d ed., 1953 S A Schelkunoff, *Electromagnetic Waves* 1943, W R Smythe *Static and Dynamic Electricity* 2d ed 1950

Reflectometer, microwave

A form of directional coupler that is used for measuring the power flowing in both directions in a wave guide A pair of single detector couplers appropriately positioned on opposite sides of the wave guide can be used for this purpose with one detector positioned to monitor transmitted power and the other detector positioned to measure power reflected back from a discontinuity in the line Each coupler receives a constant small fraction of the energy flowing in one direction in the wave guide, the energy being extracted from the wave guide through two small holes drilled ¼ wavelength apart along the length of the guide See IMPEDANCE MEASUREMENTS, HIGH FREQUENCY [J MR]

Reflex, conditioned

A response by an animal or man to a stimulus which was inadequate to elicit the response until paired for one or more times with a different and adequate stimulus The term was first used in 1903 by the Russian physiologist I P Pavlov to denote dogs' secretion of saliva at the sight and odor of food and in the feeding environment in general, as distinct from unconditioned reflex secretion produced by the actual presence of food or acid, in the mouth While "distant" salivation had been studied intensively in Pavlov's laboratory as "psychic secretion" since 1897 and the phenomenon was observed as early as 1852, Pavlov's conception of the *conditioned reflex was a revolutionary* methodological and ontological turn in the study and analysis of animal and human behavior

By 1909 other Russian workers had extended the study of the conditioned reflex to other stimulus response combinations in dogs and humans and since that time, scientists the world over have paired almost the entire repertoire of reflexes with almost the entire repertoire of stimuli in representative species of almost all phyla and classes of the entire animal kingdom and in human subjects to yield more or less successful results Approximately 3000 experiments have been reported As a result, conditioning is now considered an established general biological or psychobiological, phenomenon In conditioning, a stimulus originally inadequate to elicit some reflex (a to be continued stimulus *or a CS∼) becomes adequate (a conditioned* stimulus or a *CS*) by virtue of being applied one or more times together with the stimulus adequate to elicit the reflex (the unconditioned stimulus or the *US*) The resulting reflex becomes a conditioned reflex (a *CR*) as distinct from the unconditioned reflex to the originally adequate stimulus (the *UR*) An argument can be made for broadening the definition of conditioning by replacing the term "reflex" with that of "response" This seems logical in view of reported successful conditioning of brain waves and tropisms on the one hand, and of mental sets (a state of preparedness for making *a response to* an impending stimulus) and meanings on the other

Factors in conditioning The universality of the conditioning phenomenon does not mean that mere

ius-response the attitudinal and attentive perceptual determinants the type of reflex used, and the phyletic position (that is, fish cat, monkey or human) of the subject must be considered

The factors of intensity the *US UR* must not be too weak and the *CS* must not be too strong Optimum conditioning seems to require optimum *US UR/CS* intensity ratios

The temporal relationships conditioning proceeds best when the *CS* is applied shortly before the *US*, and is poorest when the sequence is reversed or the interval between the *CS* and *US* is prolonged

Attitudinal and attentive-perceptual determinants these operate particularly, but by no means exclusively, in the conditioning of cognitively or instructionally controllable reflexes in human subjects

Type of reflex used, and phyletic position of subject the galvanic skin reflex is readily conditioned, *but the iridic reflex is conditioned only with* great difficulty, it takes, as a rule, scores or even hundreds of trials to condition a lower invertebrate to an electric shock, but a monkey or young child *may well effect it* in one trial The relationship is, however, not a simple one Russian experiments show, for instance that while dogs salivate copiously at the sight of food, cats do not, and Kinsey's reports indicate that the sex reactions of the human female are much less conditionable to visual and verbal stimuli than are those of the human male

Properties of conditioning. Four cardinal and general properties of conditioning are (1) extinction, (2) generalization and differentiation, (3) dynamic stereotypy (Pavlov's term) or configural conditioning (G Razran's concept), and (4) higher order conditioning Extinction is the gradual diminution and final disappearance of a *CR* upon repeated applications of the *CS* without the *US* Extinguished *CR*s recover partially after a lapse of time (spontaneous recovery) and may also recover by the application of an extra stimulus (disinhibition) but finally, after an appropriate number of extinction sessions, they disappear completely and irrecoverably As a contrast to the traditional law of use, extinction is, in a way, Pavlov's most unique single contribution to both the empirics and the theory of learning (*see* LEARNING THEORIES) Generalization and differentiation cover the fact that in the initial stages of training *CR*s are evoked in some degree also by stimuli that are in some way related to their conditioned stimuli but that in later stages the *CR*s acquire a large degree of specificity This aspect of conditioning though still largely in a state of controversy *with respect to underlying mechanisms* and basic facts has been used extensively as an explanatory principle of transfer of training and of transference and related psychoanalytic concepts *See* PSYCHOANALYSIS.

Dynamic stereotypy or configural conditioning relates to results of conditioning to compound *CS*s, for example, the ringing of a bell, and the flashing of a light, a mechanical stimulation and an olfactory stimulation, combined simultaneously or in close succession In the course of such conditio- ing, the *CR* eventually comes to be elicited only l the compound *CS* and not by the individual cor ponents This property of the *CR* offers an objecti method of tracing *developmentally the acquisitic* course of stimulus patternization and stimulus se in man and animals Finally, higher order conditio ing deals with the formation of *CR*s on the basis (old *CR*s rather than *UR*s forming a *CR* of tl second order, CR^2, by combining its to-be-cond tioned stimulus, $CS^2\sim$, with a conditioned stimul of the first order, a *CS*, or forming a third-ord *CR*, CR^3, by combining its $CS^3\sim$ with a CS^2, an so on According to some schools of thought, suc conditioning offers a simple and far reaching a count of the existence and efficacy of man's cu tural needs and values *Unfortunately*, higher orde conditioning is as yet far from a wholly establishe fact in laboratory experiments with animals, an also with men, when symbolic and inferential aid are excluded

A somewhat dramatic behavioral change brough out in conditioning studies is experimental neurc sis, a laboratory induced disturbance parallelin symptoms of natural psychopathological states (*se* NEUROSIS) The main methods of producing sucl neuroses have been transforming alimentary *CR* into defensive ones and vice versa, close differentia conditioning for example, combining one stimulu repeatedly with a *UR* and another closely relate stimulus without the *UR*, long and unexpected de lays between the application of the *CS* and the *US* and general *CR* overtraining Changes in *CR* regimes, laboratory induced sleep and sedativ and stimulant drugs have been studied extensively as therapeutic agents in Russian *CR* laboratories

Neural mechanisms. *Pavlov considered the* conditioned reflex as the method for studying higher nervous (cortical) activity and formulated upon the basis of its diverse actions a complex variety of hypothetical cortical mechanisms These are internal inhibition interplay and collision of inhibition with excitation, special irradiation and concentration of the two processes transformation of one into the other induction of one by the other, protective inhibition cortical exhaustion, and the like These constitute what might be called Pavlov's conceptual cortical system criticized by some, but defended by others Recently however, extensive series of studies involving direct observations of what actually goes on in the nervous systems during conditioning have been launched in a number of *Russian laboratories*. These involve the use of rather advanced electroencephalographic histoneural biochemical and neuropharmacological techniques These direct studies provide results which are claimed to support Pavlov's main postulates, for example existence of internal inhibition and its correlation with the slowing up of brain waves its interplay with excitation generalization and differentiation of neural events and above all morphological changes in cortical stellar cells as a

result of conditioning Verification of the striking

Pavlovian, or classical conditioning, is characterized first by a strengthening of an existing response rather than by the formation of a new response and second, by an invariant occurrence of the strengthened response, the *CR*, before the strengthening one, the *UR* Examples are (1) a rat learning to press a lever more frequently when the pressing is followed by the delivery of a pellet of food or by the removal of some annoying stimulus, as shown in B F Skinner's most extensive studies, (2) a dog lifting his paw more readily when the lifting is followed by feeding (experimental work begun by J Konorsky and S Miller and since studied extensively in Russian and Polish laboratories), (3) a child pressing a rubber bulb more readily when candy is thereby obtained (numerous experiments by A G Ivanov Smolensky and his followers in the Soviet Union), (4) an animal running more speedily to a box in which he is fed, and (5) an animal, conditioned to withdraw its paw at the sound of a bell through a bell shock combination, withdrawing the paw more readily when the shock

was elicited as a response to a specific environmental stimulus rather than emitted by the animal operating on the environment.

Instrumental operant conditioning is similar to

requiring for its maintenance markedly fewer *CR* *UR* combinations The two also differ with respect to the very nature of the learned modification and the essential conditions for its occurrence Moreover there is the very important consideration that, compared with the classical conditioning, the instrumental-operant variety is effective only with certain classes of *CR*s and *UR*s and is inoperative in lower invertebrates It is more readily abolished by decortication and is seemingly different in its electroencephalographic correlates Whatever the case may be for calling instrumental operant, or reward learning "conditioning" all the evidence and logic seem to suggest that it is a more complex and phyletically more recent—thus less universal, more efficient, more central and likely more cognitive—form of learned modification than is classical Pavlovian conditioning Hence it is obviously to the latter rather than the former that we must accord the status of the simplest and most ultimate learning unit. A view that no analysis of learning is complete and ultimate without a base analysis of classical conditioning is surely well

[G R]

Marquis,

Reflex, unconditioned

A stimulus evoked, neurally mediated behavior pattern which is not subject to voluntary control and is not established by prior experience, training, or conditioning Unconditioned reflexes constitute the basis of many automatic regulatory reactions which, with machinelike precision, adapt the animal body to environmental changes Temperature regulation cardiac acceleration during exercise, certain postural adjustments, and withdrawal from a harmful stimulus are a few examples of many important reflex patterns found in mammals (*see* CARDIOVASCULAR SYSTEM, THERMOREGULATION) To the physician the reflex performance of the patient is diagnostically important All reflexes depend on neural connections between sense organs and the central nervous system (afferent path) on the one hand and between the central nervous system and the effector organs (efferent path) on the other Consequently disease involving injury to the afferent path the central nervous system or the efferent path is betrayed by aberrations of reflex performance Also because different reflexes are mediated by different neural pathways through the nervous system the distribution of reflex aberration often points to the locus of the causative lesion

Reflex arc The reflex arc is the anatomical substrate of the unconditioned reflex It consists of a chain of neurons at least one of which has its cell body located within the central nervous system Each neuron in the chain is an anatomically and metabolically discrete cell, functional linkages of neurons into chains occur when the processes of one neuron make contact with the cell body or processes of another the contact sites are called synapses *The simplest reflex chain consists of only two* neurons and hence only one synapse between sense organ and effector This monosynaptic arc is shown diagrammatically in Fig 1 The first neuron in the

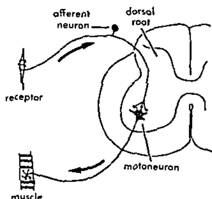


Fig 1 Diagram of monosynaptic arc

chain is called the primary afferent neuron, its cell body is situated in the dorsal root ganglion and its peripherally directed process ends in a sense organ whereas its centrally directed process enters the spinal cord to make synaptic *connection* with the second neuron which is called a motoneuron The axon of the motoneuron leaves the spinal cord by way of the ventral spinal root and runs through the peripheral nerve to reach the muscle

Many reflex arcs are more complex and have one or *more centrally located* neurons known as interneurons *or internuncial* neurons inserted (intercalated) between the primary afferent neuron and the motoneuron Such reflex arcs are multisynaptic arcs (Fig 2)

The sequence of events in monosynaptic reflex action is as follows (1) an adequate stimulus to *the sense organ* generates an action potential in the primary afferent *fiber* (2) this action potential on reaching *the intraspinal* terminals of the primary afferent fiber causes the liberation of a chemical agent which diffuses *across the synaptic* space and generates an *impulse in the* motoneuron, (3) this impulse is *propagated over the efferent* fiber to its ending in the *muscle* where the neuromuscular *transmitter agent* acetylcholine is liberated, (4) acetylcholine generates an impulse in the muscle fiber, and (5) the muscle fiber impulse sweeps over the muscle fiber *and activates the* contractile mechanism *See* ACETYLCHOLINE, MUSCLE

From the foregoing simplified description it *might be inferred* that reflex action is stereotyped *and invariant* On the contrary reflex patterns are *both variable* and adaptive The mechanism of adaptive reflex behavior is clarified by a more detailed consideration *of the central* connections of afferent paths and motoneurons

Convergence and divergence Histological examination *of* motoneurons reveals that each motoneuron *is* supplied with many synaptic terminals The terminals have the form of small (about 1μ in diameter) knobs closely applied to the dendrites and cell bodies of the motoneuron so that about 40–50% of the *somadendritic membrane* is encrusted *with synaptic knobs* The many knobs on a *single motoneuron* derive from many parent afferent fibers The motoneuron thus constitutes a final

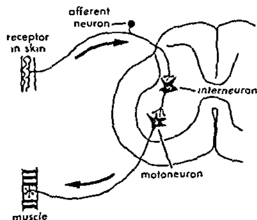


Fig 2 Diagram of multisynaptic arc

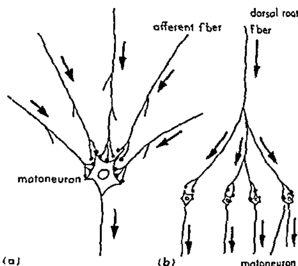


Fig 3 Diagram of convergence and divergence of afferent fibers and motoneurons (*a*) Convergence (*b*) Divergence

common path upon which many afferent fibers converge (Fig 3*a*) There is reason to suppose that many knobs must be activated within a brief period of time to initiate an impulse in the motoneuron *Firing thus results* from the nearly synchronous *activity of many afferent* fibers converging on the motoneuron, it is doubtful that activity in a single afferent fiber is sufficient to cause motoneuron discharge

Considered from the afferent side the key feature of organization is divergence Each dorsal root fiber breaks into many branches which establish connection with many *motoneurons (Fig 3b)* Thus although no single *afferent fiber alone* fires a motoneuron each fiber contributes to the excitation of many motoneurons

Excitation and inhibition Activation of afferent channels does not always excite motoneurons Some

chemical transmitter agents which they secrete

Each motoneuron is thus the recipient of a constant barrage of impulses some of which tend to activate the cell and generate an impulse others of which tend to stabilize the cell and render it quiescent At any time the behavior of the cell depends on the balance between these two opposing influences

Reciprocal innervation The distribution of excitatory and inhibitory knobs to motoneurons is not haphazard but instead is organized in such a way as to provide for unopposed reflex muscular

inhibitory knobs to the motoneurons innervating antagonists of the muscle Such reciprocal innervation provides for sensitive and delicate regulation of reflex action because it not only permits appropriate muscular contraction but also *curtails* the

tion of muscles which oppose the appropriate movement.

Reflex pathway classification Reflex arcs are often classified according to the anatomical level of their afferent and efferent connections with the central nervous system. Segmental reflexes have afferent inputs by way of the spinal dorsal roots and efferent outputs over spinal ventral roots of the same or adjacent segments. Intersegmental reflexes also have spinal dorsal roots as the afferent path but the efferent path is the ventral root of a distant spinal segment. The connection between input and output is made by intersegmental chains of inter

roots, the connection being established by long pathways descending from the brain stem to the spinal levels.

Segmental reflexes This section discusses the flexion, stepping, stretch, and lengthening reflexes.

Flexion reflex This reflex is elicited by noxious stimulation (pinching, burning, and cutting), particularly of the skin of the arm or leg, although similar stimulation of deep structures of the flayed limb is also effective. The reflex response consists of contraction of the ipsilateral flexor muscles at all joints so that the whole limb is withdrawn from the noxious stimulus. At the same time the extensor muscles relax so that withdrawal is not impeded. The afferent paths of the flexion reflex thus make reciprocal connections exciting flexor motoneurons and inhibitory extensor motoneurons. Because the adequate stimulus for eliciting the reflex is harmful

Reflex withdrawal of the stimulated limb is often associated with contraction of extensor muscles and relaxation of flexor muscles in the corresponding contralateral limb. The contralateral component is known as the crossed extension reflex but should not be considered to be a separate reflex because it is accessory to, or part of, the flexion reflex. The afferent fibers subserving the flexion reflex send to the opposite side of the spinal cord collateral branches which make reciprocal connections opposite to those in the ipsilateral spinal cord. This arrangement is known as double reciprocal innervation. The crossed extension component serves to support the weight of the body when the ipsilateral limb flexes.

The receptor for the nociceptive flexion reflex

...es and for the fact that cutaneous stimulation is particularly effective. The flexion reflex arc is multisynaptic, that is, one or more interneurons are interposed between the primary afferent neuron and the motoneurons. Interneurons tend to diffuse activity through several segments of the spinal cord

so that the resultant motoneuron discharge causes integrated muscle contraction at all joints of the limb.

Step

touch

spray

reciprocal central connections identical with those of nociceptive afferent fibers, that is, the nerve fibers are ipsilaterally (on the same side) excitatory to flexor and inhibitory to extensor motoneurons but contralaterally (on the opposite side) excitatory to extensor and inhibitory to flexor motoneurons (see POSTURE REGULATION OF). There is no reason to believe that the reflex pattern thus elicited is protective or nociceptive; instead, it probably is involved in reflex stepping or walking. Both the touch pressure receptors of the feet and the stretch sensitive secondary spindle endings are in a position to be excited alternately by normal stepping, and their afferent fibers make doubly reciprocal connections which provide for rhythmic alternate contraction of the limb muscle on the two sides. Rhythmic stepping movements can be induced artificially by concurrent electrical stimulation of nerves containing touch pressure and flower spray afferent fibers.

Stretch or myotatic reflex When a muscle is passively stretched, reflex contraction occurs opposing elongation. At the same time the antagonists of the stretched muscle relax because of inhibition of their motoneurons. The receptor for this stretch

than those mediating the flexion reflex. The stretch reflex efferent discharges are distributed only to the muscle being stretched. Similarly, inhibition is confined to the motoneurons supplying the direct (that is, acting at the same point) antagonists of the stretched muscle. The precise and discrete distribution of the stretch reflex discharge is correlated with the fact that the stretch reflex arc is monosynaptic: the afferent fibers connect directly with motoneurons.

The stretch reflex is the basis of standing. In standing the extensor muscles of the legs are stretched by the gravitational tendency of the limbs to flex at the joints. The stretch thus created excites the spindle endings and induces reflex contraction of the extensor muscles opposing gravity and maintaining upright standing. Significantly, stretch reflexes of extensor or antigravity muscles are particularly well developed.

Lengthening reflex The myotatic reflex just described causes a muscle to resist elongation. If, however, the muscle is lengthened forcefully, it first resists but then suddenly relaxes: the tension drops precipitously and the muscle may then be stretched without opposition. This phenomenon, called the lengthening reaction, is the result of a reflex mediated by stretch sensitive receptors in the tendons (Golgi tendon organs) innervated by large-diame

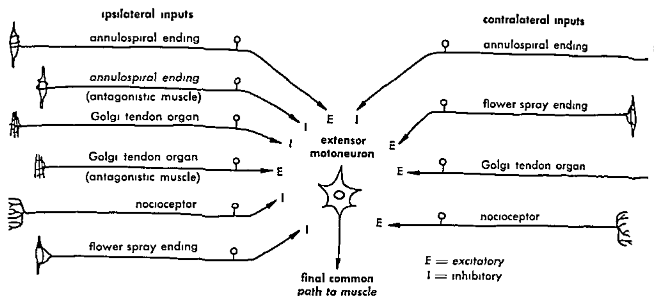


Fig 4 Diagram of a few segmental afferent pathways converging on an extensor motoneuron

ter afferent fibers The central connections of these lengthening reflex afferent fibers are the reverse of those mediating the myotatic reflex, impulses originating in Golgi tendon receptors inhibit the motoneurons supplying the stretched muscle and excite the motoneurons supplying its antagonist For this reason the reflex is often called the inverse myotatic reflex Strictly, however, the lengthening reflex is not a simple mirror image of the stretch reflex because the former traverses a three neuron (disynaptic) arc and is far less discrete in its efferent distribution than the myotatic reflex Golgi tendon organ discharge inhibits the motoneurons supplying the stretched muscle as well as those supplying other muscles at the same and adjacent joints

Golgi tendon organs, although stretch sensitive, have higher thresholds than muscle spindle endings With moderate muscle stretch, such as that imposed by gravity on extensor muscles during standing, the low threshold spindle endings dominate the motoneurons and upright posture is maintained With stronger extending forces in the face of which further opposing contraction might damage muscle and tendon a sufficient number of tendon organs discharge to overcome the excitatory barrage from spindle endings, the motoneurons are inhibited and the muscle lengthens without opposing the extending force The inverse myotatic reflex thus serves to protect the muscle against injurious overloading

Intersegmental reflexes Intersegmental reflexes serve to coordinate the reaction of the fore and hind limbs and the position of the head with those of the limbs A flexion reflex and its crossed extension component elicited in one forelimb is often accompanied by reverse changes in the hind limb, that is, ipsilateral extension and contralateral flexion Movements of the head which stretch the neck muscle are accompanied by variations in the discharge of motoneurons supplying fore and hind limb muscle Intersegmental reflexes traverse multineuronal chains and are therefore more variable than the simpler segmental reflexes At each synapse in a chain, the behavior of the postsynaptic element depends on the summated inputs, excitatory and inhibitory, converging upon it, hence, the longer the chain is, the more variable will be the performance

Suprasegmental reflexes These constitute a group of complex reflexes which serve to integrate the body and limb musculature with fixed positions or movements of the head Position movement and acceleration of the head registered by visual, labyrinthine and vestibular receptors alter discharge patterns of motoneurons controlling limb and body musculature Like the intersegmental reflexes, suprasegmental reflexes employ complex multineuronal channels and are therefore flexible and variable

Summary. Figure 4 shows a highly simplified diagram of a few of the segmental afferent pathways converging on a typical extensor motoneuron that is one element of the many which comprise the final common pathway to an extensor muscle Even ignoring the multisynaptic organization of most of the segmental inputs and the many multisynaptic intersegmental and suprasegmental inputs it can be seen that the motoneuron population is subject to a host of subtly variable afferent drives, some antagonistic and some reinforcing Each motoneuron in the population integrates the messages which impinge upon it

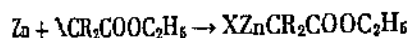
This arrangement makes the reflex arc a highly flexible mechanism in which shifts in the intensity and source of afferent bombardment may alter drastically the participation of different motoneurons and their subservient muscles in reflex action and thus give rise to an almost infinite variety of behavioral patterns *See* MUSCLE (BIOPHYSICS), NERVOUS SYSTEM, PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL, REFLEX CONDITIONED

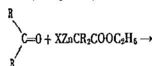
[H D P , T C R]

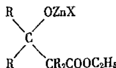
Reformatsky reaction

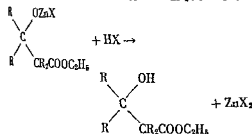
A reaction which takes place between a carbonyl compound, such as an aldehyde or ketone, and an α halo ester in the presen I zinc Hy

drolysis of the reaction mixture with dilute acid yields a β hydroxy ester The reaction is thought to proceed via an organozinc derivative, analogous to the Grignard reagent formed by the interaction of the α halo ester with the zinc The organozinc compound then adds to the carbonyl group of the aldehyde or ketone (R = alkyl, aryl, or hydrogen, X = iodine or bromine)









-formatsky reaction is a valuable tool in organic chemistry as a method for preparing β hydroxy esters and corresponding unsaturated and saturated esters and acids It also serves as a useful means of lengthening the carbon chain *See* GRIGNARD REACTION, ORGANOMETALLIC COMPOUND

[M D R]

Bibliography R Adams (ed), *Organic Reactions* vol 1 1942

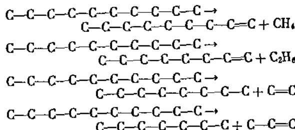
Reforming (petroleum refining)

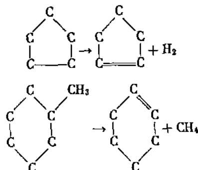
A process used for upgrading gasoline by improving its antiknock characteristics Reforming reactions occur by either thermal or catalytic methods

Thermal reforming The following reactions are typical of thermal reforming Only the carbon skeletons of the hydrocarbons are shown

Paraffins primarily undergo end of the chain cracking to yield methane, ethane, and propane as well as unsaturated gases, high boiling paraffins, and olefins. There is very little skeletal isomerization during thermal reforming Naphthenic hydrocarbons undergo dehydrogenation and some cracking to form cycloolefins There is essentially no dehydrogenation of naphthenes to aromatics and little conversion of the natural aromatics

The thermal reforming operation is usually performed at temperatures of about 1025°F and pressures of 750-1000 psig. Residence time in the reactor is about 20-40 sec. In a typical operation a





Pennsylvania straight run gasoline of 44 octane number may be reformed to give a product with an octane number of 80 (Research Clear) with a yield of 66% Higher octane number products can be obtained by thermal reforming, but the yields of gasoline from straight run stocks become uneconomically low Since catalytic reforming can produce higher quality products with higher yields the thermal reforming operations have been largely replaced by catalytic reforming In January 1958 the thermal reforming capacity in the United States was approximately 270,000 barrels per day, and the gasoline so produced accounted for about 5% of the total gasoline production

Catalytic reforming. The reactions shown on the next page are typical of catalytic reforming

The reactions are dehydrogenation of both five- and six membered ring cycloparaffins, hydrocracking (hydrogenolysis) of paraffins to form lower boiling paraffins along with propane and butanes, cyclization of paraffins to aromatics and isomerization of paraffins to more branched structures

The product from catalytic reforming consists of aromatic and paraffin hydrocarbons, along with a small quantity of unconverted naphthenes The aromatics are concentrated in the higher boiling fraction while the paraffins constitute the bulk of the lower boiling fraction

Catalytic reforming operations commonly employ a supported platinum catalyst A proper balance between the dehydrogenation and acid acting components is built into the catalyst, so that it removes hydrogen rearranges the carbon skeleton, and splits the molecule

The catalytic reforming step is performed at temperatures of 800–1050°F and pressures of 200–1000 psig An important feature of catalytic reforming is the recycling of a hydrogen rich gas that acts to suppress those side reactions which tend to form carbonaceous deposits on the catalyst The usual amount of hydrogen recycle is 3–15 moles of hydrogen per mole of charging stock

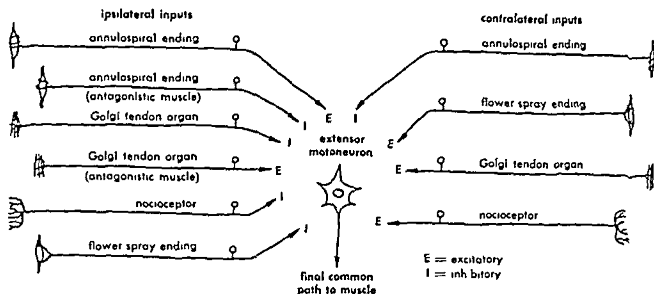


Fig 4 Diagram of a few segmental afferent pathways converging on an extensor motoneuron

ter afferent fibers The central connections of these lengthening reflex afferent fibers are the reverse of those mediating the myotatic reflex, impulses originating in Golgi tendon receptors inhibit the motoneurons supplying the stretched muscle and excite the motoneurons supplying its antagonist For this reason the reflex is often called the inverse myotatic reflex Strictly, however the lengthening reflex is not a simple mirror image of the stretch reflex because the former traverses a three-neuron (disynaptic) arc and is far less discrete in its efferent distribution than the myotatic reflex Golgi tendon organ discharge inhibits the motoneurons supplying the stretched muscle as well as those supplying other muscles at the same and adjacent joints

Golgi tendon organs, although stretch sensitive have higher thresholds than muscle spindle endings With moderate muscle stretch such as that imposed by gravity on extensor muscles during standing, the low threshold spindle endings dominate the motoneurons and upright posture is maintained With stronger extending forces in the face of which further opposing contraction might damage muscle and tendon a sufficient number of tendon organs discharge to overcome the excitatory barrage from spindle endings, the motoneurons are inhibited, and the muscle lengthens without opposing the extending force The inverse myotatic reflex thus serves to protect the muscle against injurious overloading

Intersegmental reflexes. Intersegmental reflexes serve to coordinate the reaction of the fore and hind limbs and the position of the head with those of the limbs A flexion reflex and its crossed extension component elicited in one forelimb is often accompanied by reverse changes in the hind limb that is, ipsilateral extension and contralateral flexion Movements of the head which stretch the neck muscle are accompanied by variations in the discharge of motoneurons supplying fore and hind limb muscle Intersegmental reflexes traverse multineuronal chains and are therefore more variable than the simpler segmental reflexes At each synapse in a chain, the behavior of the postsynaptic element depends on the summated inputs, excitatory and inhibitory, converging upon it, hence, the longer the chain is the more variable will be the performance

Suprasegmental reflexes. These constitute a group of complex reflexes which serve to integrate the body and limb musculature with fixed positions or movements of the head Position movement and acceleration of the head registered by visual, labyrinthine and vestibular receptors alter discharge patterns of motoneurons controlling limb and body musculature Like the intersegmental reflexes, suprasegmental reflexes employ complex multineuronal channels and are therefore flexible and variable

Summary. Figure 4 shows a highly simplified diagram of a few of the segmental afferent pathways converging on a typical extensor motoneuron that is one element of the many which comprise the final common pathway to an extensor muscle Even ignoring the multisynaptic organization of most of the segmental inputs and the many multisynaptic intersegmental and suprasegmental inputs, it can be seen that the motoneuron population is subject to a host of subtly variable afferent drives, some antagonistic and some reinforcing Each motoneuron in the population integrates the messages which impinge upon it

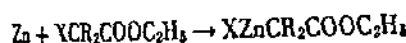
This arrangement makes the reflex arc a highly flexible mechanism in which shifts in the intensity and source of afferent bombardment may alter drastically the participation of different motoneurons and their subservient muscles in reflex action and thus give rise to an almost infinite variety of behavioral patterns *See* Muscle (biophysics), Nervous system, Psychology, physiological and experimental, Reflex, conditioned

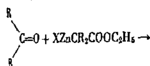
[H D P , T C R]

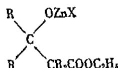
Reformatsky reaction

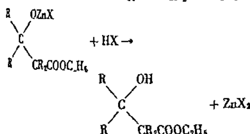
A reaction which takes place between a carbonyl compound, such as an aldehyde or ketone, and an α halo ester in the presenc zinc Hy

nlysis of the reaction mixture with dilute acid yields a β hydroxy ester The reaction is thought to proceed via an organozinc derivative analogous to the Grignard reagent formed by the interaction of the α halo ester with the zinc The organozinc compound then adds to the carbonyl group of the aldehyde or ketone (R = alkyl aryl or hydrogen X = iodine or bromine)









The use of zinc in the Reformatsky reaction has

the more reactive Grignard reagents Hence the Reformatsky reaction is a valuable tool in organic chemistry as a method for preparing β hydroxy esters and corresponding unsaturated and saturated esters and acids It also serves as a useful means of lengthening the carbon chain *See* GRIGNARD REACTION ORGANOMETALLIC COMPOUND

[M D R]

Bibliography R Adams (ed) *Organic Reactions*, vol. 1 1942.

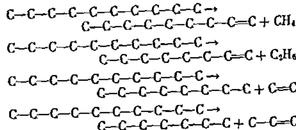
Reforming (petroleum refining)

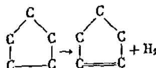
A process used for upgrading gasoline by improving its anti knock characteristics Reforming reactions occur by either thermal or catalytic methods

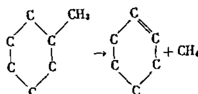
Thermal reforming The following reactions are typical of thermal reforming Only the carbon skeletons of the hydrocarbons are shown

Paraffins primarily undergo end-of the-chain cracking to yield methane ethane and propane as well as unsaturated gases high boiling paraffins and olefins. There is very little skeletal isomerization during thermal reforming Naphthenic hydrocarbons undergo dehydrogenation and some cracking to form cycloolefins There is essentially no dehydrogenation of naphthenes to aromatics and little conversion of the natural aromatics

The thermal reforming operation is usually performed at temperatures of about 1025°F and pressures of 50-1000 psig Residence time in the reactor is about 20-40 ... In a typical operation a







Pennsylvania straight run gasoline of 44 octane number may be reformed to give a product with an octane number of 80 (Research Clear) with a yield of 66% Higher octane number products can be obtained by thermal reforming, but the yields of gasoline from straight run stocks become uneconomically low Since catalytic reforming can produce higher quality products with higher yields the thermal reforming operations have been largely replaced by catalytic reforming In January 1958 the thermal reforming capacity in the United States

and
5%

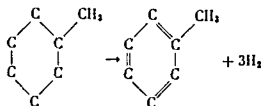
Catalytic reforming The reactions shown on the next page are typical of catalytic reforming

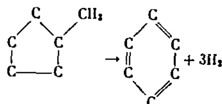
The reactions are dehydrogenation of both five and six membered ring cycloparaffins hydrocracking (hydrogenolysis) of paraffins to form lower boiling paraffins along with propane and butanes cyclization of paraffins to aromatics and isomerization of paraffins to more branched structures

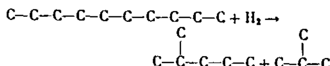
The product from catalytic reforming consists of aromatic and paraffin hydrocarbons along with a small quantity of unconverted naphthenes The aromatics are concentrated in the higher boiling fraction while the paraffins constitute the bulk of the lower boiling fraction

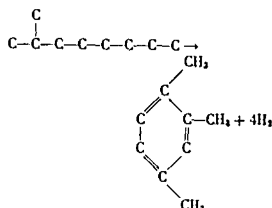
Catalytic reforming operations commonly employ a supported platinum catalyst A proper balance between the dehydrogenation and acid acting components is built into the catalyst so that it removes hydrogen rearranges the carbon skeleton and splits the molecule

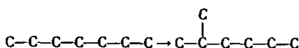
The catalytic reforming step is performed at temperatures of 800-1050°F and pressures of 200-1000 psig An important feature of catalytic reforming is the recycling of a hydrogen rich gas that acts to suppress those side reactions which tend to form carbonaceous deposits on the catalyst. The usual amount of hydrogen recycle is 3-15 moles of hydrogen per mole of charging stock. This hydrogen











partial pressure is sufficient to maintain a clean catalyst surface for long periods of time from 6 months to 2 years for example When the catalyst becomes fouled with carbonaceous deposits and other impurities such as metals it is replaced In some installations the carbonaceous deposits are burned off and the catalyst is returned to service until it is finally deactivated for reasons other than carbon deposition

The use of a platinum containing catalyst for the reforming of straight run gasolines has undergone a *rapid acceptance by the petroleum industry since* the first unit was installed in 1949 The catalytic *reforming capacity in the United States in 1958* was in excess of 1 500 000 barrels per day In 1957 the gasoline produced by catalytic reforming amounted to 36% of the total gasoline production

Catalytic reforming units usually have a fraction ation section where the fresh feed is distilled This removes as overhead the lower boiling components of the feed such as hexane and lower boiling hy drocarbons, and rejects the material boiling above the gasoline range (400+°F) The C_7 400°F frac tion is mixed with hydrogen, preheated to the reac tion temperature, and the reforming carried out in three to four reactors in series Intermediate reheat ing is necessary since the *over all heat of reaction* is endothermic The effluent from the reactors goes through heat *exchangers to a* separator where the liquid product is separated and sent to the stabilizer to produce a *finished gasoline as the bottoms* prod uct and to remove, as overhead propane and a part of the butane *produced during the reaction* Part of the gas component removed from the separator consisting mostly *of hydrogen, is withdrawn from* the system, the remainder is recycled to admix with the fresh feed

Many of the installations made since 1954 include a pretreatment section wherein *the sulfur* com pounds and other impurities are removed by means of a hydrotreating step utilizing, usually, a cobalt molybdena alumina catalyst In earlier installations, scrubbers were installed on the recycle gas in or der to reduce the sulfur concentration entering the reactors The reduction in sulfur makes it possible to use more severe conditions of temperature and pressure thereby producing gasolines of higher oc tane rating The pretreating step also removes from the feed arsenic compounds which otherwise poison the platinum catalyst *See* ALKYLATION AROMATIZATION, CRACKING, DEHYDROGENATION, HYDROCARBON, ISOMERIZATION, PETROLEUM PROC ESSING [V H]

Refraction (molar)

A physical constant which is dependent upon molecular structure and which is defined by the equation

$$R_M = \frac{n_r^2 - 1}{n_r^2 + 2} \quad \frac{M}{\rho}$$

In this expression n is the refractive index of the material in question M and ρ are the molecular weight and density respectively

A light wave may be considered as a rapidly *alternating electric and magnetic field* As it passes from air or a vacuum into any material substance it *interacts with the electrons of this medium* This induces an oscillatory motion in the electrons The light wave in turn suffers a decrease in velocity The refractive index is a measure of this interac tion and depends upon the polarizability of the molecule The molar refraction then is a function of the number kinds and arrangement of atoms in a molecule According to the simplest picture the molar refraction of a substance is equal to the actual volume occupied by the molecules in one mole Some typical values determined at 20°C with the *D* line of sodium as a light source are listed *in* Table 1

To a good approximation each atom and each structural characteristic such as a double bond or

Table 1 Molar refractions of some compounds at 20 C

Compound	R_M cm³/mole
Carbon tetrachloride, CCl_4	26 51
Benzene, C_6H_6	26 18
Ethanol C_2H_5OH	12 78
Water H_2O	3 75

Table 2. Molar refractions of molecular units

Group	R_M, cm^3/mole
CH_2	4 618
H	1 100
C	2 418
Double bond,	
C=C	1 733
Cl	5 967
−C≡N	5 459
Br	8 865
O (hydroxyl)	1 525

r linkage, make a definite contribution to the ar refraction For this reason, molar refraction il value in characterizing substances of similar iposition Several such group or atomic contributions are listed in Table 2

f the wavelength of the light used in the measurement of the refractive index is sufficiently long *l the substance has no permanent dipole moment*, the Maxwell relation $n_r^2 = \epsilon$ is applicable, i the dielectric constant of the substance Under se conditions, the molar refraction is equal to : molar polarization *See* POLARIZATION (DIELECTRICS), REFRACTION OF WAVES [F J J]

efraction of waves

he change of direction of propagation of any wave tenomenon which occurs when the wave velocity hanges The *term is most frequently applied to* isible light but it also applies to all other electromagnetic waves, as well as to sound and water waves.

The physical basis for refraction may be readily understood with the aid of Fig 1 Consider a suc-

. . .

. . . .

wn the velocity of propagation is faster than n their original medium, they are refracted away from the normal

Snell's law. The simple mathematical relation governing refraction is known as Snell's law If waves traveling through *a medium* at speed v_1 are incident on a boundary surface at angle θ_1 (with respect to the normal), and after refraction enter the second medium at angle θ_2 (with the normal) while traveling at speed v_2, then

$$\frac{v_1}{v_2} = \frac{\sin\theta_1}{\sin\theta_2}$$

The index of refraction *n* of a medium is defined as the ratio of the speed of waves in vacuum *c* to their speed in the medium. *Thus* $c = n_1v_1 = n_2v_2$ and therefore

$$n_1 \sin\theta_1 = n_2 \sin\theta_2 \qquad (1)$$

Fig 1 Physical basis for Snell's law

The refracted ray the normal to the surface, and the incident ray always lie in the same plane

The relative index of refraction of medium 2 with respect to that of medium 1 may be defined as $n = n_1/n_2$ Snell's law then becomes

$$\sin\theta_1 = n \sin\theta_2 \qquad (2)$$

For sound and other elastic waves which require a medium in which to propagate, only this last form has meaning Equation (2) is frequently used for light when one medium is air, whose index of refraction is very nearly unity

When the wave travels from a region of low velocity (high index) to one of high velocity (low index) refraction occurs only if $(n_1/n_2)\sin\theta_1 \leqq 1$ If θ_1 is too large for this relation to hold, one has $\sin\theta_2 > 1$, which is meaningless In this case the waves are totally reflected from the surface back into the first medium The largest value that θ_1 can have without total internal reflection taking place is known as the critical angle, θ_c Thus $\sin\theta_c = n_2/n_1$ When the angle of incidence $\theta_1 < \theta_c$, refraction occurs as in Fig 2*a* When $\theta_1 = \theta_c$, the *emergent ray just grazes the* surface (Fig 2*b*) Total internal reflection (Fig 2*c*) represents the only practical case for which 100% of the incident energy is reflected and none is absorbed When it is desired to change the direction of a beam of light without loss of energy, totally reflecting prisms are often used, as in prism binoculars

If waves travel through a medium having a continuously varying index of refraction, the rays follow smooth curves with no abrupt changes of direction Suppose (Fig 3) that $n = n(y)$, and that the incident ray lies in the xy plane If θ is the angle between the direction of the ray and the y axis,

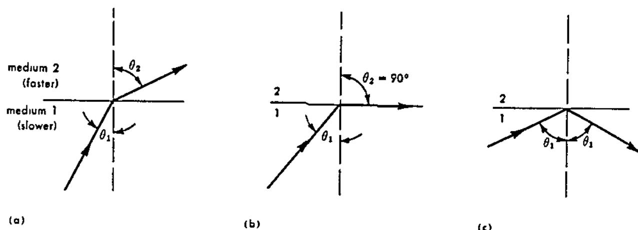


Fig 2 Behavior of ray traveling from medium of high refractive index to medium of low index (a) $\theta_1 < \theta_c$, ray is refracted (b) $\theta_1 = \theta_c$, ray grazes surface (c) $\theta_1 > \theta_c$, ray is reflected

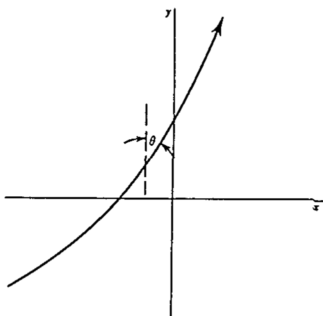


Fig 3 Path of light in medium having continuously varying index of refraction, $n = n(y)$

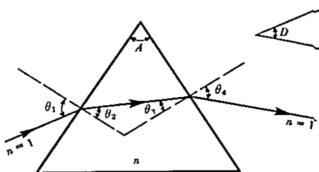


Fig 4 Refraction of light by a prism

then Snell's law can be written in the differential form

$$\frac{d\theta}{dn} = -\frac{1}{n}\tan\theta \qquad (3)$$

In a particular case Eq (3) can be integrated to give the path of the ray

Visible light Many cases of refraction that are of interest occur for visible light The refraction of light by a prism in air affords a particularly simple and useful example

As a ray passes through the prism of Fig 4 its total deflection or deviation $D = \theta_1 + \theta_4 - A$, where A is the vertex angle of the prism Also, by Snell's law,

$$n = \frac{\sin\theta_1}{\sin\theta_2} = \frac{\sin\theta_4}{\sin\theta_3}$$

It is found that the deviation is a minimum when the ray passes through the prism symmetrically (that is, when $\theta_1 = \theta_4$) For minimum deviation,

$$n = \frac{\sin \frac{1}{2}(A + D)}{\sin \frac{1}{2} A} \qquad (4)$$

For a given prism the dispersion, or lateral spread of the spectrum formed, is maximum for that wavelength of light which passes through the prism at minimum deviation *See* PRISM, OPTICAL

For most optical materials, the dispersion $dn/d\lambda$ (λ is the wavelength) is negative so that red light is bent less than blue light Typical values of n for optical materials range from 1 5 for ordinary crown glass, 1 7 or 1 8 for dense flint glass, up to 2 42 for diamond For water, n is 1 33 Some special substances have even higher values Many substances show anisotropy in the refraction of light, with different indices of refraction in different directions *See* OPTICAL MATERIALS

For a lens (Fig 5) refraction occurs at both surfaces If the lens is thin and the rays all make small angles with the axis of the system, application of Snell's law to the two spherical surfaces yields the well known lens formula

$$\frac{1}{s} + \frac{1}{s'} = \frac{1}{f} \qquad (5)$$

where s is the object distance from the lens s' the image distance and f the focal length of the lens *Magnifying instruments such as binoculars* telescopes microscopes, and projectors make use of refraction by lenses or prisms in their operation *See* LENS OPTICAL

Double refraction Some anisotropic single crystals such as those of calcite and quartz are bire-

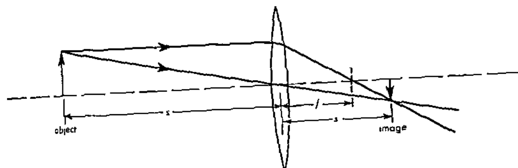


Fig. 5 Refraction of light by a lens

fringent, or doubly refracting If one looks through such a crystal at a dot on a piece of paper he sees two images As the crystal is rotated in the plane of the paper one image remains stationary while the other appears to rotate about it

Two separate rays propagate through the crystal they are called the ordinary ray and the extraor

The extraordinary ray does not propagate ndicularly to its wavefronts The separation ;en the two rays depends upon the direction ıich the light travels through the crystal rela o that of the crystal s optic axis Light travel parallel to the optic axis is only singly re ıed.

refringent crystals are either uniaxial or bi il, depending upon whether they have one optic or two They are said to be positive or nega , depending upon whether the velocity of prop tion (within the crystal) of the extraordinary t is greater or less than that of the ordinary re. Calcite is a uniaxial negative crystal quartz uniaxial positive crystal The most commonly ed biaxial crystal is mica Doubly refracting tals are frequently employed as polarizers ch as the Nicol prism *See* CRYSTAL OPTICS OLARIZED LIGHT

Refractometry The measurement of indices of refraction, called refractometry can be made in

Alternatively the critical angle for total internal reflection may be measured Another method is to observe visually the apparent thickness of a slab of material by looking straight through it, and to compare this with the real thickness as measured with a micrometer Then

$$n = \frac{\text{real thickness}}{\text{apparent thickness}}$$

Interferometric methods are particularly convenient for gases. In the Jamin refractometer for example, a simple count of fringes as the gas is slowly admitted to an initially evacuated tube in the optical path yields n These techniques can also be used for solids particularly when the material is available in the form of thin films *See* INTERFEROMETRY

Refractometry is an important tool in analytical chemistry For example information about the composition of an unknown solution can frequently be obtained by measurement of its index of refraction *See* REFRACTOMETRIC ANALYSIS

Atmospheric refraction Gases have indices of refraction only slightly greater than unity In general $n - 1$ is proportional to the density of the gas or to the ratio of pressure to absolute temperature The index of refraction of the earth's atmosphere increases continuously from 1 000000

than they appear to be The effect decreases from a maximum of about 35 min of arc for an object on the horizon to zero at the zenith where the light enters the atmosphere at perpendicular incidence Thus the sun (and all other bodies) appears to rise 2 or more min earlier (depending upon latitude) and to set 2 or more min later than would be the case without refraction This must be taken into account when the altitude of a celestial body is observed for navigational purposes

Other manifestations of atmospheric refraction are the mirages and 'looming of distant objects which occur over oceans or deserts, where the vertical density gradient of the air is quite uniform over a large area The twinkling of stars is caused by the rapid small fluctuations in density along the light path in the atmosphere Rainbows are produced by the multiple reflections refraction and dispersion of sunlight by spherical raindrops *See* METEOROLOGICAL OPTICS MIRAGE, RAINBOW, TWINKLING STARS

Other electromagnetic waves Although refraction is most frequently encountered for the visible portion of the spectrum it is of importance for other electromagnetic radiation For very long wavelength radiation the index of refraction of many materials is equal to the square root of the dielectric constant k In general $dn/d\lambda$ is negative except in the regions of so-called anomalous dis-

persion near absorption bands On the short wave length side of an absorption band n can be less than 1 00 Since it is the phase velocity of the wave rather than the group velocity which is involved in the definition of the index of refraction this does not represent a violation of the principle of relativity, that is that energy cannot be propagated at a velocity faster than the velocity of light in vacuum (*see* RELATIVITY *see also* GROUP VELOCITY PHASE VELOCITY) At very high frequencies the index of refraction of all materials is also slightly less than unity *See* ABSORPTION (ELECTROMAGNETIC RADIATION)

Refraction plays a role in the propagation beyond the line of sight of radio waves in the earth s atmosphere (*see* RADIO WAVE PROPAGATION)

The interaction of electromagnetic radiation with more or less opaque substances is often described in terms of a complex index of refraction The real part of this quantity has the usual meaning for the small amount of light which penetrates into the material before it is absorbed The imaginary part is a measure of the absorption

Sound waves The velocity of sound in a gas is proportional to the square root of the absolute temperature Because of the vertical temperature gradients in the atmosphere refraction of sound can be quite pronounced As in mirage formation to allow large scale refraction the temperature at a given height must be uniform over a rather large horizontal area If the temperature decreases with altitude (the usual situation) sound waves initially traveling at a small angle with the horizontal are refracted upward A sound out of doors is thus not normally audible at a great distance However if there is a temperature inversion (as over a body of water on a calm sunny day) the waves would be refracted downward This is the main reason that sound carries long distances across water on a calm day On a windy day the horizontal temperature strata are broken up and the sound is dissipated

Refraction accompanied by reflection accounts for the fact that large explosions are sometimes heard in several distinct regions at surprisingly large distances with zones of silence in between A temperature inversion at high levels can refract the waves downward into a zone of audibility The sound is then reflected from the ground and must again be refracted downward to give the next zone of audibility *See* SOUND WAVE MOTION IN FLUIDS

Seismic waves The velocity of elastic waves in a solid depends upon the modulus of elasticity and upon the density of the material Waves propagating through the solid earth are refracted by changes of material or changes of density World wide observations of earthquake waves enable scientists to draw conclusions on the distribution of density within the earth These waves may be totally internally reflected at the boundary of the core It was through such observations that the existence of the much denser core of the earth was first postulated

Refraction of compressional waves from explosions set off on the ground is (combined with reflection) used in prospecting for oil natural gas and minerals which have large differences in density and elastic constants from the surrounding rock *See* SEISMOLOGY

Water waves The speed of water waves in shallow water is proportional to the square root of the depth As the waves enter shallower water they travel more slowly As a train of waves approaches a coastline obliquely its direction of travel becomes more nearly perpendicular to the shore because of refraction *See* SHORE PROCESSES *See also* OPTICS, GEOMETRICAL, WAVE MOTION IN LIQUIDS

[J W ST

Bibliography M Born and E. Wolf *Principles of Optics* 1959, F A Jenkins and H E White *Fundamentals of Optics*, 3d ed 1957

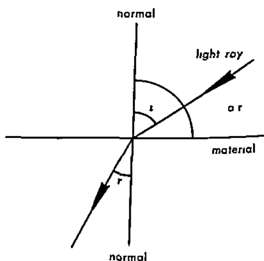
Refractometric analysis

A method of chemical analysis based on the measurement of the index of refraction of a substance When light impinges on the surface of a material at an angle i to the normal to the surface, its direction is changed on passing into the material so that it then travels at an angle r to the normal The index of refraction is defined as follows

$$\text{Index of refraction} = \frac{\sin i}{\sin r} = n_{\mathrm{D}}^{20}$$

It varies as a function of temperature and wave length of light and also of pressure in gases Refractive indices are usually measured at 20°C using the yellow D line of the sodium spectrum The indices of refraction of a few substances are water 1 333 benzene 1 5014 chloroform 1 4464 and acetone 1 3589

The most common type of refractometer is the Abbe refractometer It is simple to use requiring but a drop or two of sample and allowing a measurement of refractive index to be made in a minute or two with a precision of 0 0001 More precise measurements of refractive indices within 0 00003



Refraction of light

s are made with an interferometer based on the terference of light Its precision is 0 000001

The measurement of refractive index is used to entify compounds whose other physical constants e quite similar Because minute amounts of impurities often cause a measurable change in the refractive index of a pure material refractive index often used as a criterion for purity A measurement of refractive index gives information as to the gross amount of impurity it does not serve to identify the impurity In order to give qualitative information on measurements would have to be made at different wavelengths a rare procedure In systems containing only two components such as water and alcohol mixtures or aqueous salt solutions refractive index is a sensitive and rapid method of determining the composition See OPTICAL METHODS OF CHEMICAL ANALYSIS REFRACTION OF WAVES

[R F G]

Refractory

One of a number of ceramic materials for use in high temperature structures or equipment The term high temperatures is somewhat indefinite but usually means above about 1000°C or temperature or pyrometric cone equivalent (PCE) the number of the standard pyrometric cone which deforms under heat treatment in the same manner as the fireclay Thus the minimum PCEs for low intermediate high and super duty fireclays are 19 29 31/32 and 33 respectively (See PYROMETRIC CONE) Fireclays are also classified by their working properties into two other classes plastic (those which form a moldable mass when mixed with water) and flint (a hard rocklike clay that does not become plastic when mixed with water) In general flint clays have higher PCEs than plastic clays and are mixed with the latter to form the higher grades of firebrick See CLAY COMMERCIAL

High alumina refractories are made from clays which contain in addition to the alumina (Al_2O_3) in the clay minerals hydrates of aluminum oxide and which are often loosely referred to as bauxite These raise the total Al_2O_3 content and make the material more refractory Different grades are distinguished on the basis of the total Al_2O_3 content (50 60 70% alumina refractories)

Sillimanite and kyanite are anhydrous alumino silicate minerals ($Al_2O_3 \cdot SiO_2$) used to make special refractory objects such as crucibles tubes and muffles or used as an addition to fireclay to control its shrinkage during firing

Silica refractories These the second largest

tantalum are used

The biggest use of refractories is in the steel industry for construction of linings of equipment such as blast furnaces, hot stoves and open hearth furnaces. Other important uses of refractories are for cement kilns glass tanks nonferrous metallurgical furnaces ceramic kilns steam boilers, and paper plants. Special types of refractories are used in rockets, jets, and nuclear power plants Many refractory materials such as aluminum oxide and silicon carbide are also very hard and are used as abrasives some applications, for example aircraft brake linings make use of both character

... magnesite dolomite or chrome ore termed basic refractories (because of their chemical behavior) and (4) a miscellaneous category usually referred to as special refractories See METAL ...

Aluminosilicate refractories Fireclay is the ... material from which the bulk (about 70%) of refractories is manufactured. Different grades are distinguished according to the softening tempera

added to assist in bonding both before and after firing The quality of silica refractories is to a great extent determined by the amount of Al_2O_3 impurity even small amounts having a deleterious effect on refractoriness This is just opposite to the case of alumina in fireclays where a higher alumina content means greater refractoriness High grade silica brick contains less than 0 6% Al_2O_3 and even the standard grade contains less than 1% During firing the mineral quartz transforms to cristobalite and tridymite the high temperature forms of silica Since these are less dense than quartz, the brick expands on firing and the true density of the solid is often taken as a test of adequate firing An example would be the case in which the density of acceptably fired material must be below 2 35 g/cm^3 The outstanding characteristic of silica is its ability to withstand high loads at elevated temperatures for example as a sprung arch roof 30 or 40 ft wide over an open hearth The hearth may be operated within 50°C of the melting point of silica See SILICON

Semisilica refractories are made from clay with a high silica (sand) content (over 70% total silica) their main advantage is their dimensional stability when heated or fired Apparently the expansion of the silica as sand offsets the contraction of the clay

Basic refractories Magnesite refractories are so named because magnesium carbonate mineral was for many years the sole raw material Since World

War II, sea water has become a significant source of magnesium oxide refractory and such material is often called sea water magnesite. In any case, the raw material is calcined to *form a material largely* magnesium oxide, MgO, about 5% iron oxide is usually added before calcining *See* MAGNESITE

Chrome refractories are made from chrome ore, a complex mineral containing oxides of chromium, iron, magnesium, aluminum, and other oxides crystallized in the spinel structure These crystals are usually embedded in a less refractory matrix called gangue *See* CHROMITE

In an attempt to combine the best properties of each, magnesite and chrome are often mixed to *form chrome-magnesite* or magnesite chrome refractories *(the first named is the dominant constituent)*

Dolomite is a mixed calcium magnesium carbonate, $CaMg(CO_3)_2$, which, when calcined to a mixture of MgO *and* CaO, *is used in granular form to* patch the bottoms of open hearths and also to make bricks

Miscellaneous materials. Special refractories are made of a great many materials, and it is possible to mention here only a few of the more important

Silicon carbide (SiC) is used for many refractory shapes, its outstanding properties being good thermal and electrical conductivity (it is used to make electric heating elements for furnaces), good heat shock resistance, strength at high temperatures, and abrasion resistance The first silicon carbide refractories were bonded with clay so that the refractory properties of the bond placed the ultimate limit on the material A method of making self *bonded silicon carbide has been developed to re-*move this limitation Although silicon carbide tends to oxidize to form SiO_2 and either CO or CO_2, the silica oxidation product forms a glassy coating on the remaining material and to a certain extent protects it from further oxidation

Insulating firebrick is made from refractory clays to which a combustible material (sawdust cork *coal)* has been added, when this burns during the firing operation, it leaves a brick of high porosity The low thermal conductivity of insulating brick reduces heat losses from furnaces and the low bulk density and consequent low heat capacity reduce the amount of heat needed to bring the furnace itself up to temperature The main disadvantage of such bricks is their low strength, but even this is useful in that they can be cut or ground to shape quite readily

Pure oxides, of which alumina (Al_2O_3) is the prime example, are used for many special refractories Some, such as beryllia (BeO) thoria (ThO_2), and uranium oxide (UO_2), are of particular interest for nuclear applications

Carbides, nitrides, borides, silicides, and sulfides of various sorts have been considered as refractory materials and some study made of them, aside from a few carbides and nitrides, however, none have found much use

Cermets are an intimate mixture of a metal a *nonmetal, for example,* Al_2O_3 and chromium though the nonmetal may be an oxide, it is n commonly a carbide or *nitride (as in cemen* tungsten carbide) *See* CERMET.

Carbon, generally in the form of graphite used for such *equipment as crucibles and as s* per nozzles in ladles for steel casting A potenti *very large use of carbon is in* blocks for constr *tion of blast furnace* hearths Graphite has v good thermal shock resistance and moderate e *trical* conductivity, does not melt but rather s *limes at a significant* rate only at temperatu well above 3000°C, is quite inert chemically, an wet by very few molten materials Its main dis vantage, common to all nonoxide materials at h temperatures, is that it oxidizes, since the produ are all gaseous, they offer no protection agai further oxidation *See* GRAPHITE

Manufacture of refractories. *Standard cera* techniques are used (*see* CERAMIC TECHNOLOG Hand molding, once widely used, is used today o for special shapes and small orders The extrus or stiff mud process is used for plastic firecla very often the extruded blanks are repressed hydraulically rammed to form special shapes, example, T sections of refractory pipe Power pre ing of simple shapes is the most widely used for ing method Hot pressing and hydrostatic pressi are used for some special refractories Slip casti is used for special refractory shapes Fusion ca ing is commonly used for glass tank blocks, th are either mainly Al_2O_3 or Al_2O_3 with significa amounts of SiO_2, ZrO_2 or both

Refractories are generally fired in tunnel kil *but some periodic kilns are still used, particula* for special shapes *See* KILN

Some types of basic refractories, known as che ically bonded are pressed with a chemical bind such as magnesium oxychloride and installed wi out firing Some of these, the steel clad refra tories are encased in a metal sheath at the time pressing When the refractory is heated after i stallation the iron oxidizes and reacts with t refractory forming a tight bond between the i dividual bricks

In all refractory products and in *unfired brick* particular the maximum possible formed densi is desired To this end careful crushing and sizin of raw materials are carried out so that, as far a possible the gaps between large pieces are fille with smaller particles and the space between the with still smaller and so on In the case of clay re fractories it is customary to use prefired (ca cined) clay or crushed fired rejects (both ar known as grog) to increase the density and to re duce the firing shrinkage

Properties of refractories *A high melting poin* is of course necessary in a refractory but man other properties must be considered in choosing a refractory for a specific application

A definite melting point is characteristic of pur materials, actual minerals from which refractorie

are made for example clay are far from pure and hence do not melt at a specific temperature Rather, they form increasing amounts of liquid as the temperature is increased above a certain minimum temperature at which liquid first appears This characteristic of gradual softening is specified by the PCE of the material

High temperature strength is important for refractories, but most materials become plastic and

A knowledge of the thermal expansion of high temperature materials is important first so that allowance can be made in furnace construction (long tunnel kilns must be built with expansion joints of several inches every 10 ft or so) and second because of its relation to thermal shock resistance.

Thermal conductivity determines the amount of heat that will flow through a furnace wall under given conditions and a knowledge of this property is essential to furnace design

Thermal shock resistance is the ability of a specimen to withstand without cracking a difference in temperature between one part and another For example, if a red hot brick is dropped into cold water it is likely to shatter since the outside cools and contracts while the center is still hot This cracking is often referred to as thermal spalling the term spalling meaning any cracking off of large pieces of brick Other causes of spalling are mechanical (hitting the brick and knocking off a piece) and structural (a reaction in the brick which changes the mineral structure and causes cracking) Thermal shock resistance is enhanced by high strength low Young s modulus low thermal expansion, and sometimes, depending on conditions high thermal conductivity Whether or not a given specimen cracks under heat shock depends not only on the material of which it is made but also on its size and shape and on the test conditions, for example whether it is dropped into water or into still air at the same temperature

Chemical properties of various kind are of im

... be as low as possible Turning to high temperature chemistry the rate of corrosion of refractories by molten slags and iron oxide fumes is vital to the length of service rendered Reference to the appropriate phase equilibrium diagrams may give some indication of which combinations of slag and refractory will react but in most applications actual tests are needed to make any precise predictions. The rate of corrosion depends to a great extent on such physical factors as the porosity of the refractory and whether or not it is wet by the slag.

Carbon deposition is another chemical reaction which affects the life of refractories The reaction ... with the refractory but is catalyzed by substances in it When carbon monoxide, perhaps in the top of a blast furnace comes in contact with certain iron compounds which can occur in fire clays its reduction to carbon is catalyzed This carbon deposits at the site of the catalyst in the brick and causes the brick to shatter The effect is most pronounced around 500°C, much below this temperature, the rate of reaction is too slow, and much above it, the equilibrium oxygen pressure necessary for the reduction is lower than is found in practice Although the reaction is not completely understood it has been found that high temperature firing of the fireclay refractories converts the iron to a form which does not catalyze the carbon deposition

The bursting of spinel (chrome) refractories in contact with iron oxide is another high temperature chemical reaction it is not thoroughly understood but appears to be related to oxidation and reduction reactions in the refractory [M C M]

Bibliography J H Chesters *Steelplant Refractories*, 2d ed rev 1957, F H Norton *Refractories*, 3d ed 1949

Refrigerated truck

A type of insulated truck (or trailer) equipped with a means for keeping the interior cool for the hauling of fresh perishables or below freezing for the hauling of frozen products High temperature truck bodies have 3–4 in of light density batt or board insulation for transporting fresh commodities at 35°F to 40°F Low temperature truck bodies have 6 in of insulation for transporting frozen products at 0°F or lower temperature Mechanical refrigeration water ice dry ice (solid CO_2) or eutectic holdover plates may be used for cooling The holdover plates filled with brine are cooled down at night at a central refrigeration plant Heating often must be provided for winter operation *See* DRY ICE REFRIGERATION [H M HE]

Refrigeration

The cooling of a space or substance below the environmental temperature The art was known to the ancient Egyptians and people of India who used evaporation to cool liquids in porous earthen jars exposed to dry night air and to early Chinese Greeks and Romans who used natural ice or snow stored in underground pits for cooling wine and other delicacies In the late eighteenth and early nineteenth centuries natural ice cut from lakes and ponds in winter was stored underground for use in summer The technique of mechanical refrigeration began with the invention of machines for making artificial ice Great strides have been made in the past 50 years in the application of mechanical refrigeration to fields other than ice making including the direct cooling and freezing of perishable foods and air conditioning for industry and human comfort

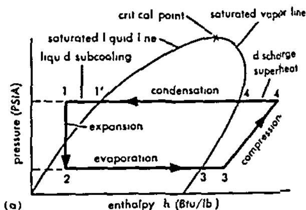
Mechanical refrigeration is primarily an application of thermodynamics wherein the cooling medium or refrigerant goes through a cycle so that

it can be recovered for reuse The commonly used basic cycles, in order of importance, are vapor compression, absorption, steam jet or steam ejector, and air Each cycle operates between two pressure levels, and all except the air cycle use a two-phase working medium which alternates cyclically between the liquid and vapor phases

Vapor compression cycle. The vapor compression cycle (Fig 1) consists of an evaporator in which the liquid refrigerant boils at low temperature to produce cooling, a compressor to raise the pressure and temperature of the gaseous refrigerant, a condenser in which the refrigerant discharges its heat to the environment, usually a receiver for storing the liquid condensed in the condenser, and an expansion valve through which the liquid expands from the high pressure level in the condenser to the low pressure level in the evaporator This cycle may also be used for heating if the useful work is taken off at the condenser level instead of at the evaporator level (*see* HEAT PUMP)

The theoretical vapor compression cycle can best be analyzed on the pressure-enthalpy or temperature entropy coordinates for a two-phase fluid (Fig 2) Enthalpy is a parameter replacing heat content It equals internal energy plus the product of the pressure and the volume divided by 778 and is expressed in units of Btu per lb Entropy is a parameter obtained by dividing the heat flow by the average absolute temperature during the change It is expressed in units of Btu per lb per degree Rankine (°R = °F + 460) Process 1–2 represents adiabatic (constant enthalpy) expansion, 2–3′ constant temperature (and pressure) evaporation, 3′–3 suction superheating at constant pressure, 3–4, ideal frictionless adiabatic (constant entropy) compression, 4–4′, removal of discharge superheat at constant pressure, 4′–1′ condensation at constant pressure (and temperature), and 1′–1 liquid subcooling at constant pressure

The efficiency of a heat power cycle is defined as the ratio of useful output to energy input. For a heat engine efficiency is less than unity Efficiency



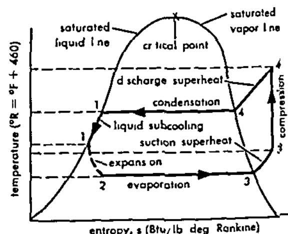


Fig 2 Vapor-compression cycle shown on (a) pressure-enthalpy diagram and (b) temperature-entropy diagram

is not very meaningful for the refrigeration and heat pump cycles, where instead the term coefficient of performance (*CP*) is used *Referring to* the theoretical cycle (Fig 2), the refrigeration CP is the ratio of cooling effect in evaporator 2–3 to compressor energy input 3–4 and the heat pump CP is the ratio of heating effect in condenser 4–1 to compressor energy input 3–4 The coefficient of performance may be considerably greater than unity and the theoretical heat pump CP is 1 plus refrig

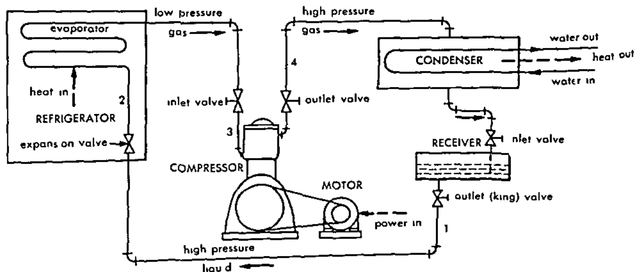


Fig 1 Vapor compression cycle

eration CP For theoretical cycles operating be tween the same temperature levels, the heat pump CP is the reciprocal of the heat engine efficiency

Absorption cycle. The absorption cycle ac complishes compression by using a secondary fluid to absorb the refrigerant gas which leaves the evaporator at low temperature and pressure Heat is applied by means such as steam or gas flame, to distill the refrigerant at high temperature and pres sure. The most used refrigerant in the basic cycle (Fig 3) is ammonia, the secondary fluid is then water The condenser, receiver expansion valve, and evaporator are essentially the same as in any vapor-compression cycle The compressor is re placed by an absorber, generator, pump heat ex changer, and reducing valve

The operation of the cycle is based on the prin ciple that the vapor pressure of a refrigerant is lowered by the addition of an absorbent having a lower vapor pressure, and the greater the quantity of absorbent used the more the depression of the vapor pressure of the refrigerant (*see* DALTON'S LAW) By maintaining the solution in the absorber at the proper temperature and concentration the vapor pressure of the solution can be kept lower than that of the refrigerant in the evaporator Spraying the weak solution in the absorber will then cause the refrigerant vapor to flow from the evaporator to the absorber The strong solution thus formed in the absorber is then pumped through a heat exchanger to the generator, where heat is applied to release the refrigerant vapor Then fol lows condensation expansion and evaporation as in the standard vapor compression cycle Except for small units, an indirect system is used wherein

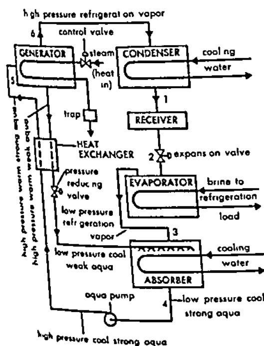


Fig 3. Basic absorption cycle

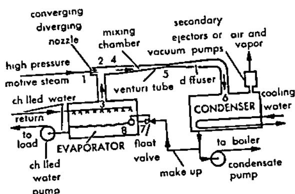


Fig 4 Steam jet water vapor cycle

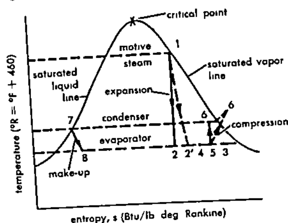


Fig 5 Temperature-entropy diagram (steam ejector cycle)

brine is cooled and circulated to the actual refrig eration load

For air conditioning, water is the refrigerant and lithium bromide is the absorbent In terms of the basic absorption cycle (Fig 3), from 1 to 2 the high pressure liquid refrigerant is expanded into the evaporator where brine is usually cooled from 2 to 3 the low pressure refrigerant vapor is drawn into the absorber, from 3 to 4 the low pressure refrigerant vapor is absorbed in the weak solution, f 4 to 5 the low pressure strong solution is

the condenser The hot weak solution drains back to the absorber through the heat exchanger and pressure-reducing valve

Steam jet cycle. The steam jet cycle uses water as the refrigerant High velocity steam jets provide a high vacuum in the evaporator causing the water to boil at low temperature and at the same time compressing the flashed vapor up to the condenser pressure level Its use is limited to air conditioning and other applications for temperatures above 32°F

The basic steam jet or ejector cycle (Fig 4) is usually analyzed on temperature-entropy coordi nates (Fig 5) High pressure motive steam (Fig 4) at 1 is expanded to a low absolute pressure

it can be recovered for reuse The commonly used basic cycles, in order of importance, are vapor-compression, absorption, steam jet or steam ejector, and air Each cycle operates between two pressure levels, and all except the air cycle use a two-phase working medium which alternates cyclically between the liquid and vapor phases

Vapor compression cycle. The vapor compression cycle (Fig 1) consists of an evaporator in which the liquid refrigerant boils at low temperature to produce cooling, a compressor to raise the pressure and temperature of the gaseous refrigerant, a condenser in which the refrigerant discharges its heat to the environment, usually a receiver for storing the liquid condensed in the condenser, and an expansion valve through which the liquid expands from the high pressure level in the condenser to the low pressure level in the evaporator This cycle may also be used for heating if the useful work is taken off at the condenser level instead of at the evaporator level (*see* HEAT PUMP)

The theoretical vapor compression cycle can best be analyzed on the pressure enthalpy or temperature entropy coordinates for a two phase fluid (Fig 2) Enthalpy is a parameter replacing heat content It equals internal energy plus the product of the pressure and the volume divided by 778 and is expressed in units of Btu per lb Entropy is a parameter obtained by dividing the heat flow by the average absolute temperature during the change It is expressed in units of Btu per lb per degree Rankine (°R = °F + 460) Process 1–2 represents adiabatic (constant enthalpy) expansion, 2–3′, constant temperature (and pressure) evaporation, 3′–3, suction superheating at constant pressure, 3–4, ideal frictionless adiabatic (constant entropy) compression, 4–4′ removal of discharge superheat at constant pressure, 4′–1′, condensation at constant pressure (and temperature), and 1′–1, liquid subcooling at constant pressure

The efficiency of a heat power cycle is defined as the ratio of useful output to energy input For a heat engine, efficiency is less than unity Efficiency

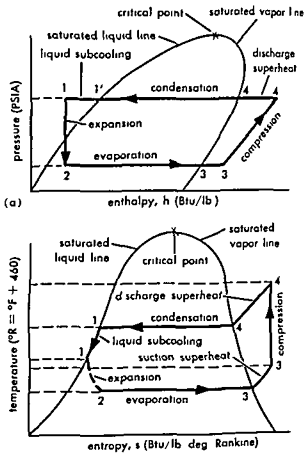


Fig 2 Vapor-compression cycle shown on (a) pressure-enthalpy diagram and (b) temperature-entropy diagram

is not very meaningful for the refrigeration and heat pump cycles, where instead the term coefficient of performance (CP) is used Referring to the theoretical cycle (Fig 2), the refrigeration CP is the ratio of cooling effect in evaporator 2–3 to compressor energy input 3–4 and the heat pump CP is the ratio of heating effect in condenser 4–1 to compressor energy input 3–4 The coefficient of performance may be considerably greater than unity and the theoretical heat pump CP is 1 plus refrig

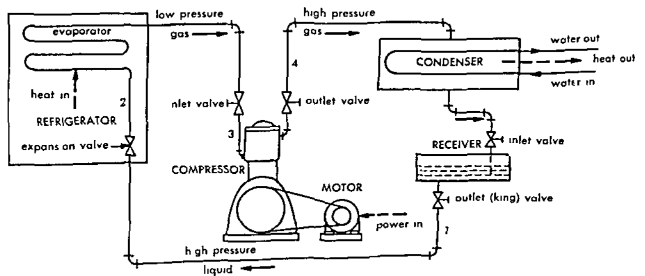


Fig 1 Vapor compression cycle

able of common refrigerants

Refrigerant trade name	ASRE† std refrigerant number	Chemical formula	Boiling point at atm press in °F
Air	729		−318
Ammonia	717	NH_3	−28
Carbon dioxide*	744	CO_2	−109 (sublimes)
Freon-11	11	CCl_3F	74 8
Freon-12	12	CCl_2F_2	−21 6
Freon-21	21	$CHCl_2F$	48 1
Freon 22	22	$CHClF_2$	−41 4
Freon 114	114	$C_2Cl_2F_4$	38 4
Methyl chloride*	40	CH_3CL	−10 8
Methylene chloride*	30	CH_2Cl_2	105 2
Sulfur dioxide*	764	SO_2	14
[illegible]	[illegible]	[illegible]	212

for industrial refrigeration Freon 11 and Freon 12 for commercial and air conditioning work where nontoxic refrigerants are necessary A secondary cooling liquid that does not change from the liquid phase is called a brine Solutions of sodium chlo-

tials, or of the centrifugal type for low pressure differentials. Early ammonia compressors were horizontal double-acting slow speed units built like steam engines Modern reciprocating compressors are vertical single-acting multi-cylinder high speed units built like automobile engines Ammonia and other large compressors require water jacketing whereas most Freon compressor cylinders are air cooled

Condensers Refrigeration condensers may be air cooled for small and medium capacities or water cooled of the shell and tube shell and coil or double pipe types. Because of the large quantities of condensing water required cooling towers or spray ponds are commonly used to recool the water for reuse An evaporative condenser is a device combining a condensing coil and a forced draft cooling tower in a single unit

Evaporators Refrigerant evaporators are the cooling units placed in the room or fluid to be cooled. Plain pipe coils or finned coils with or without forced circulation of the fluid being cooled are commonly used Shell and tube coolers or tanks

Expansion valve The main flow control in a vapor-compression system is the expansion valve It permits the liquid refrigerant to expand from the high pressure in the condenser to the lower pressure in the evaporator The expansion causes part of the liquid to evaporate and thereby to cool the

the evaporator temperature A float

maintain a fixed liquid level in the evapo- domestic refrigerators, a capillary tube restricts flow from condenser to evaporator For complete automatic operation additional controls are required to maintain the desired temperature in the evaporator, to regulate the compressor operation and the flow of the condensing medium, and to provide safety protection *See* COLD STORAGE, COOLING TOWER, DRY ICE, ICE MANUFACTURE, MARINE REFRIGERATION, PACKING HOUSE, REFRIGERATED TRUCK, REFRIGERATOR, REFRIGERATOR CAR [H M HE]

Bibliography American Society of Heating Refrigerating and Air Conditioning Engineers, *Heating Ventilating, Air Conditioning Guide*, 1959, W Stoecker, *Refrigeration and Air Conditioning*, 1958

Refrigeration cycle

A sequence of thermodynamic processes whereby heat is withdrawn from a cold body and expelled to a hot body Theoretical thermodynamic cycles consist of nondissipative and frictionless processes (*see* THERMODYNAMIC PROCESSES) For this reason a thermodynamic cycle can be operated in the forward direction to produce mechanical power from heat energy or it can be operated in the reverse direction to produce heat energy (*see* HEAT PUMP) from mechanical power The reversed cycle is used primarily for the cooling effect that it produces during a portion of the cycle and so is called a refrigeration cycle

In the refrigeration cycle a substance, called the refrigerant is compressed cooled, and then expanded In expanding the refrigerant absorbs heat from its surroundings to provide refrigeration After the refrigerant absorbs heat from such a source the cycle is repeated Compression raises the temperature of the refrigerant above that of its natural surroundings so that it can give up its heat in a heat exchanger to a heat sink such as air or water Expansion lowers the refrigerant temperature below the temperature that is to be produced inside the cold compartment or refrigerator The sequence of processes performed by the refrigerant constitutes the refrigeration cycle When the refrigerant is compressed mechanically, the refrigerative action is called mechanical refrigeration

There are many methods by which cooling can be produced (*see* REFRIGERATION) The methods include the noncyclic melting of ice or the evaporation of volatile liquids, as in local anaesthetics, the Joule Thomson effect which is used to liquify gases the reverse Peltier effect which produces heat flow from the cold to the hot junction of a bimetallic thermocouple when an external emf is imposed and utilization of the paramagnetic effect to reach extremely low temperatures (*see* PARAMAGNETISM) However large scale refrigeration or cooling in general calls for mechanical refrigeration acting in a closed system

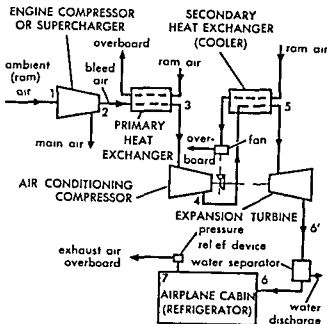


Fig 6 Open air-cycle bootstrap system for airplanes

at 2 through a converging diverging nozzle Path 1-2 (Fig 5) is the ideal expansion and 1-2′ the actual expansion allowing for nozzle friction Water vapor in the evaporator at 3 is entrained by the motive steam at 4, the latter having lost some of its energy from 2′ to 4 because the entrainment efficiency is rather low The motive steam at 4 plus the entrained moisture at 3 are forced through the venturi tube where the velocity of the incoming mixture is reduced and converted into pressure head in the condenser at 6 Path 5-6 is the ideal compression and 5-6′ the actual compression taking into account compression efficiency Typical efficiencies are nozzle 88%, entrainment 65%, and compression 80%

In the evaporator or flash chamber, part of the water is evaporated to cool the rest of the water, which is circulated to the cooling load Make-up water must be added from 7 to 8 Typical operating conditions are 100°F (2 in Hg absolute pressure) in the condenser and 40°F (0 25 in Hg abs press) chilled water in the evaporator The condensate from the condenser is pumped back to the boiler, secondary ejectors, or vacuum pumps, are required to remove the air and maintain the high vacuum Considerably more water is required for condensing than for a vapor compression system of the same capacity

Air cycle. The air cycle used primarily in airplane air conditioning, differs from the other cycles in that the working fluid, air, remains as a gas throughout the cycle Air coolers replace the condenser, and the useful cooling effect is obtained by a refrigerator instead of by an evaporator A compressor is used, but the expansion valve is replaced by an expansion engine or turbine which recovers the work of expansion Systems may be open or closed In the closed system, the refrigerant air is completely contained within the piping and components, being continuously reused In the open system, the refrigerator is replaced by the space to be cooled, the refrigerant air being expanded directly into the space rather than through a cooling coil

One of the typical open air cycle systems used on airplanes is called the "bootstrap" system (Fig 6) It may be analyzed theoretically on the temperature entropy coordinates (Fig 7) From 1 to 2, ambient air is compressed ideally in the engine or supercharger of the airplane Part of this high pressure air is bled through a primary heat exchanger where it is cooled from 2 to 3 by ram air, that is, ambient air compressed by the forward motion of the airplane From 3 to 4 this air is further increased in pressure by the compressor of the refrigeration machine, from 4 to 5 the air is cooled by ram air in the cooler or secondary heat exchanger, and from 5 to 6 the air is further cooled in the expansion turbine, ideally without moisture. However, there is entrained moisture, about 70% of which is removed by the water separator, resulting in an approximate path 5-6′ as the air is warmed by the heat given up by the condensation and removal of this moisture The balance of the moisture evaporates in the cabin and contributes to the cooling effect 6′-7, where 6′ is the dry air rated temperature of the air as it enters the cabin Dry air rated temperature is the temperature which would be attained by the expansion of the refrigerant air in the absence of any moisture condensation Refrigerant air at 7, having absorbed the heat load in the cabin, is released overboard through a pressure relief device The equipment is proportioned so that the work of expansion is recovered and is sufficient to drive the compressor and the cooler fan to approximate constant entropy expansion and compression However, compressor and turbine efficiencies and heat exchanger pressure drops, neglected in this analysis, reduce the ideal performance

Refrigerants and equipment. The working fluid in a two-phase refrigeration cycle is called a refrigerant Commonly used refrigerants are listed in the table Ammonia and Freon 22 are most important

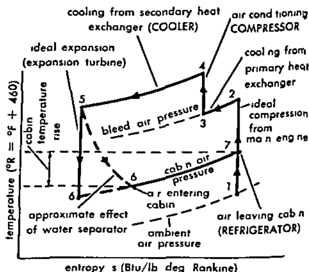
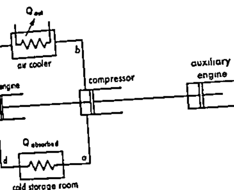


Fig 7 Temperature-entropy diagram (air-cycle bootstrap system)



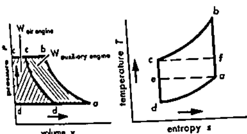


Fig. 3 Reverse Brayton cycle, or dense air refrigeration cycle

readily observable signal as it approaches and passes point e in the course of its evaporation but it would undergo a temperature rise if it accepted heat beyond point 3 This cycle using dry compression, is the one which has won overwhelming acceptance for refrigeration work

Reverse Brayton cycle The reverse Brayton cycle constitutes another possible refrigeration cycle

e refrigerators operated on the Brayton cycle

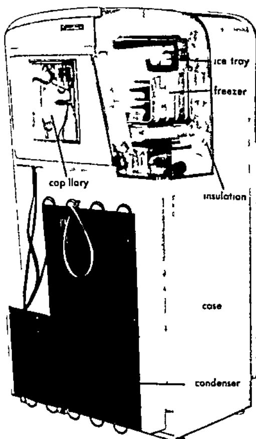
ers ble constant pressure cooling The high pressure air next expands reversibly in the engine and exhausts at low temperature The cooled air passes through the cold storage chamber it picks up heat at constant pressure and finally returns to the suc

nly way that air can reject a significant quantity of heat al ng the line *b-c* is to have *b* considerably higher than *c* Correspondingly if the cold body service temperature is *a* the air must be at a much lower temperature in order to accept heat along path *d-a* If a reverse Carnot cycle were used with a working substance undergoing changes in state the fluid would traverse path *a f c e* instead of path *a b c d* The reverse Carnot cycle would accept more heat along path *e-a* than the reverse Brayton cycle removes from the cold body along path *d a* Also, the work area required by the reverse Carnot cycle is smaller than the corresponding area for

Refrigerator

An insulated, cooled compartment If it is large enough for the entry of a man, it is a walk in box, otherwise it is a reach in refrigerator Cooling may be by mechanical or gas refrigeration, by water or dry ice, or by brine circulation Temperatures maintained depend upon the requirements of the product stored generally varying from 55°F down to 0°F and sometimes lower

A household or domestic refrigerator is a factory built self contained cabinet having a total storage space of less than 16 ft³ (as illustrated) Modern designs have a main compartment for holding food above freezing a second compartment for storage below freezing and trays for the freezing of ice cubes The cabinets are usually all metal with 2–3½ in of insulation The refrigeration unit is usually electric-motor driven but gas refrigerators motivated by the thermal energy of burning gas are used extensively in areas where cheap natu



Back of domestic refrigerator with portions of case and insulation cut away to show construction (*Philco Corporation*)

Reverse Carnot cycle. The purpose of a refrigerator is to extract as much heat from the cold body as possible with the expenditure of as little work as possible The yardstick in measuring the performance of a refrigeration cycle is the coefficient of performance, defined as the ratio of the heat removed to the work expended The coefficient of performance of the reverse Carnot cycle is the maximum obtainable for stated temperatures of source and sink Figure 1 depicts the reverse Carnot cycle on the T s plane

The appearance of the cycle in Fig 1 is the same as that of the power cycle, but the order of the cyclic processes is reversed Starting from state 1 of the figure, with the fluid at the temperature T_H of the hot body, the order of cyclic events is as follows

1 Isentropic expansion, 1–2, of the working fluid to the temperature T_c of the cold body

2 Isothermal expansion, 2–3, at the temperature T_c of the cold body during which the cold body gives up heat to the working fluid in the amount Q_c, represented by the area 2–3–*b*–*a*

3 Isentropic compression, 3–4, of the fluid to the temperature T_H of the hot body

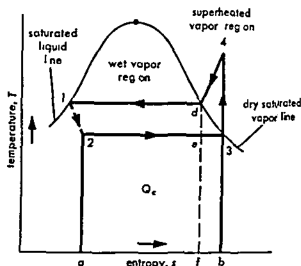
4 Isothermal compression, 4–1, at the temperature T_H of the hot body During this process, the hot body receives heat from the working fluid in the amount Q_H represented by the area 1–4–*b*–*a* The difference $Q_H - Q_c$ represented by area 1–2–3–4 is the net work which must be supplied to the cycle by external systems.

Figure 1 indicates that Q_c and the net work rectangles each have areas in proportion to their vertical heights Thus the coefficient of performance, defined as the ratio of Q_c to net work, is $T_c/(T_H - T_c)$

The reverse Carnot cycle does not lend itself to practical adaptation because it requires both an expanding engine and a compressor Nevertheless its performance is a limiting ideal to which actual refrigeration equipment can be compared

Modifications to reverse Carnot cycle. One change from the Carnot cycle which is always made in real vapor compression plants is the substitution of an expansion valve for the expansion engine Even if isentropic expansion were possible, the work delivered by the expansion engine would be very small and the irreversibilities present in any real operations would further reduce the work delivered by the expanding engine The substitution of an expansion valve or throttling orifice with constant enthalpy expansion changes the theoretical performance but little and greatly simplifies the apparatus A typical vapor compression refrigeration cycle is shown in Fig 2 it is essentially a reverse Rankine cycle The irreversible adiabatic expansion 1–2 differs only slightly from the vertical isentropic expansion

Another practical change from the ideal Carnot cycle substitutes dry compression 3–4 for wet compression *e d* in Fig 2 placing state 4 in the superheat region above ambient temperature, the process is called dry compression in contrast to the wet compression of the Carnot cycle Dry compression introduces a second irreversibility by exceeding the ambient temperature thus reducing the coefficient of performance However, dry compression is usually preferred because it simplifies the operation and control of a real machine Vapor gives no



№ 3892

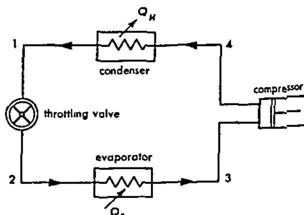


Fig 2 Vapor-compression refrigeration cycle substitutes valve for expansion engine

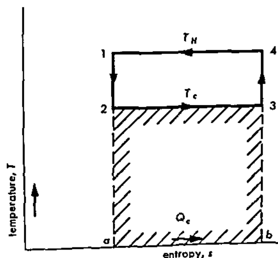


Fig 1 Reverse Carnot cycle

extensive dissociation of its component cells If a sponge is forced through fine bolting cloth, the re sulting individual cells will reassemble into a num ber of small aggregates. Each of these aggregates possesses the capacity for developing into a new sponge (Fig. 1) Ameboid movements of individual dissociated cells and the sorting-out and the reunit ing of cell types on the basis of cellular affinities play the major roles in the reaggregation and re constitution which leads to the establishment of a new individual.

Coelenterates In the invertebrates above the sponges, regenerative capacity is greatest in the coelenterates flatworms echinoderms, and arthro pods. Among the coelenterates the hydroids have been subjected to the most extensive study A fresh water hydra for example is a sessile animal which possesses a base for attachment, a long body and a group of tentacles at the upper end When a hydra is cut into two parts the basal half ordinarily re generates a new upper end of the body complete with tentacles, an isolated upper part of the body will establish a new lower portion with a basal at tachment. When a central portion of the body pos sessing neither tentacles nor basal attachment is isolated it will regenerate new basal structures at one end and tentacles at the other A fragment of the body as small as 1/200 of the original animal can reconstitute a completely new individual Compara ble results have been obtained with the more com plex marine hydroids (Figs 2 and 3)

A large amount of research on hydroids has been devoted to problems of polarity Under certain ex perimental conditions, such as altering the oxygen

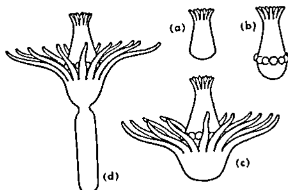


Fig 2 Reorganization in the hydroid polyp *Tubularia* (a) A small part produces the apical part of a polyp (b c d) More basal parts are formed if more material is present (*After C M Child from C P Raven An Out line of Developmental Physiology, McGraw Hill, 1954*)

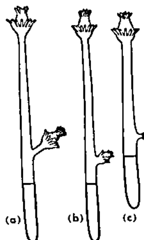


Fig 3 Organization by transplanted parts of the stem in the hydroid polyp *Corymorpha* (a) An apical part of the stem induced a complete new hydranth in 48 hours. (b) More basal parts of the stem induced smaller outgrowths in the same time (c) More basal parts of the stem inducing abnormal outgrowths at the same time (*After C M Child from C P Raven, An Outline of Developmental Physiology McGraw-Hill, 1954*)

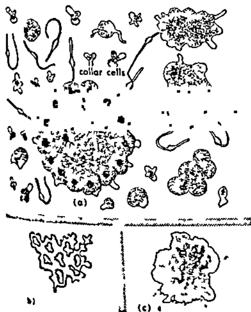


Fig. 1 (a) Appearance of *Microciona* tissue 10 minutes after being squeezed through bolting cloth (b) Reticulate reunition mass formed from such tissue (c) Later stage of the same, practically a young sponge (*After C. E. Wilson, 1911, from L H Hyman, The Invertebrates, vol 1, McGraw-Hill 1940*)

concentration at one end of the organism the type of regeneration can be changed When the cut end of the body that normally regenerates tentacles is subjected to a low concentration of oxygen it forms basal structures instead treating the cut basal end with a high concentration of oxygen results in tentacle regeneration Other experimental meth ods have produced similar results

The sea anemones representing another type of coelenterate are capable of considerable regenera tion but have been studied far less than the hy droids The regenerative capacity is least in the jellyfishes

Platyhelminthes Various species of planaria a type of flatworm have long been favored organ isms for experimental studies on regeneration An adult planaria possesses a relatively high degree of organization It has a head with eyes and a sim

ral gas is available Low temperature household re frigerators, or home freezers, for the storage of frozen foods are manufactured in both the chest and the upright, or vertical, types

A commercial refrigerator is any factory built refrigerated fixture, cabinet, or room that can read ily be assembled and disassembled, in contrast to a built in refrigerator Commercial or built in refrig erators are used in restaurants, markets, hospi tals, hotels, and schools for the storage of food and other perishables A meat cooler is a refrigerator held at about 33°F for the storage of fresh meats Refrigerators for lower temperatures down to 0°F and below are called freezer boxes Insulation thicknesses vary from 3–8 in depending upon the service In markets and stores, commercial display refrigerators may be of the self service type from which the customer helps himself Both vertical types with glass doors, which the customer opens, and chest types with open tops are used Electric refrigeration units may be built into each fixture, or remotely located *See* COLD STORAGE, REFRIG ERATION [H M HE]

Refrigerator car

An insulated railway freight car provided with cool ing equipment for transportation of perishables under controlled temperature to prevent deterioration in transit Temperatures range from 65°F for ba nanas to below 0°F for frozen products, 4–4½ in of insulation are recommended Mechanical refrig eration is fast becoming popular, but the standard method of cooling has been with ice, or ice mixed with up to 30% salt (NaCl) Ice bunkers are provided at each end of the car, with fan circulation often included Alternatively, the car can be venti lated without being cooled or heated Tariffs vary accordingly *See* REFRIGERATED TRUCK, REFRIGERA TION [H M HE]

Regeneration (biology)

The replacement by an organism of parts of the body which have been lost or severely injured The term is comprehensive and covers a wide range of restorative activities in a variety of organisms Some authors prefer to use the term reconstitution rather than regeneration

There is a long record of observations and ex periments on regeneration References to the phenomenon are found in Aristotle and Pliny The first extensive experiments on record are those of Abra ham Trembley, who studied fresh water hydras The results of this work, begun in 1740, aroused wide interest and soon led to the testing by other naturalists of the regenerative capacities of a num ber of organisms C Bonnet, 1745, was among the first to study regeneration in worms, and of par ticular interest is the work of L Spallanzani, 1768, who is credited with the first regeneration experiments on the limbs and tails of amphibians

REGENERATIVE CAPACITY

The capacity for regeneration varies greatly among different groups of organisms Among the invertebrates, many of the hydroids flatworms an nelids, echinoderms, and arthropods can replace major portions of the body In certain instances, particularly in sponges, a few cells or a small fragment of the original organism is capable of reconstituting a completely new individual In the vertebrates, the highest capacity for regeneration is found in the Amphibia, of which many species can regenerate a complete limb, a tail, portions of the eye, the lower jaw, and a number of other highly organized structures

As a rule, the structures formed as a result of regeneration are duplicates of the original struc tures and possess all of their functional charac teristics Under some circumstances, however, the regenerate may be of a different type than the origi nal structure After removal, the eye of a crustacean may be replaced by an antenna A head may be formed at the posterior end of a flatworm instead of a tail An amphibian limb may regenerate supernumerary limbs that were not previously pres ent

Many organisms, both invertebrate and verte brate although they may be incapable of regener ating complex organs or major portions of the body, have the capacity for reconstituting various types of tissues Examples are the continual or periodic replacement by various animals of skin scales feathers teeth, antlers, the lining of the alimentary canal and some components of the reproductive tract Such activities which represent a phase in the normal life cycle of an individual, are often referred to as repetitive or physiological re generation Wound healing and the repair of bone fractures can likewise be regarded as types of re generation

Among the major problems are the sources of the cells which enter into regenerative activities and the manner in which they achieve the requisite potentialities for forming specialized tissues, new organs or discrete portions of the body Also of primary importance are problems of polarity and the relationship of local and organismic factors in the establishment of growth patterns

REGENERATION IN INVERTEBRATES

The following are examples of regeneration in invertebrates

Protozoa. All of the main groups of free living protozoans exhibit some capacity for regeneration This includes replacement of a major portion of the organism as well as individual organelles Reconstitution of parts of a single cell is considerably different from regeneration of structures in a multi cellular organism and it is difficult to make com parisons Experiments have been most extensive on the ciliates, such as *Paramecium* and its relatives. Many observations indicate the importance of the macronucleus whose presence in a fragment of a protozoan is essential for regenerative activity

Porifera. Although regeneration in the usual sense of the term does not occur to any great ex tent in sponges, these organisms possess an extraor dinary capacity for reconstitution of the body after

by exposure to radium or x rays in appropriate dosage and the regenerative capacity of the worm is lost.

The segmental nature of the earthworm body and its high capacity for regeneration have provided a favorable field for quantitative studies of growth limitation. In one species of earthworm which has been extensively studied the number of segments which can be reconstituted posteriorly is a linear function of the distance in segments between the level of the cut and the anterior end of the animal. Regeneration ceases when the number of segments characteristic of the species has been restored.

Echinodermata Among the echinoderms regeneration in the various types of starfishes and brittle stars is a matter of common observation to anyone who examines these animals at the seashore. All species so far as known readily regenerate new arms. When an animal is cut into a number of pieces a new organism develops from each piece, *provided that a portion of the central disk is present*. Often severe injury to an arm results in its pinching off at the base, *a new arm then regenerates* from the remaining short stump (*see* AUTOTOMY). In some cases a number of arms may arise from a single stump to produce a highly atypical organism. Sea urchins have a far lower regenerative capacity than *starfishes but they can reconstitute* various portions of the skeleton and also the tube feet. The sea cucumbers, so far as they have been studied, have been found to possess remarkable capacities for regeneration. In some species reconstitution of the body takes place after an animal has been cut into two or three pieces. Sea cucumbers exhibit an unusual phenomenon of evisceration *when roughly handled*. Under such circumstances they are able to regenerate a completely new set of visceral organs.

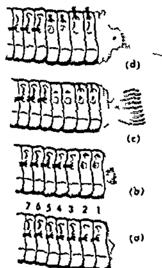


Fig. 7 Four stages (a,b,c,d) in the regeneration of the rostral end in *Sabella*. As regeneration proceeds the foremost four segments change from the abdominal to the thoracic type by the loss of their old hooks and bristles and the formation of new ones in another position. This change proceeds in a rostro-caudal succession. (After Berrill and Mees from C. P. Raven, *An Outline of Developmental Physiology*, McGraw-Hill, 1954)

Arthropoda The arthropods include many divergent forms: the various types of crustaceans, the centipedes and millipedes, the insects, scorpions, and spiders. The Crustacea, especially the lobster, crab, and crayfish, have been extensively studied with respect to regenerative capacity. The appendages of the Crustacea are the antennae, which are sensory in nature, the feeding appendages around the mouth, and the various types of legs used for locomotion, food procurement, and protection. Also there are highly complex eyes mounted on movable stalks. *All of these structures are capable* of regeneration. It is often observed that the two *large claws of a lobster are unequal in size*. When this is the case, the smaller claw is a regenerate [illegible] has not

regeneration in the Crustacea. For example, after amputation of an eye an antenna may be regenerated in its place. A nerve ganglion lies at the base of each optic stalk. If the ganglion is undisturbed *at the time an optic stalk is amputated*, a new optic stalk and eye will be regenerated. If the optic ganglion is severely injured or is removed at the time of amputation, an antenna will be regenerated instead of an eye. Other instances of atypical regeneration include the formation of a walking leg in place of one of the mouth parts, or an abdominal leg in place of a thoracic leg.

Compared with the Crustacea, regeneration in other forms of arthropods has been little studied. *Some centipedes, millipedes, and spiders can regenerate* legs. In adult insects little regeneration occurs, although in some species legs *can be re-established* in the larval and pupal stages of insects regenerative capacity is far higher than in the adult.

REGENERATION IN VERTEBRATES

Among vertebrates the amphibians possess the highest capacity for regeneration. Various species are able to regenerate limbs, parts of the eye, the *tail, the lower jaw*, and a number of other structures. Many types of fish can regenerate fins, some *portions of the gills, barbules, and scales*. In the reptiles the principal structure which is capable of regeneration is the tail of *some lizards. The tail possesses a special "breaking point"* between two vertebrae posterior to the pelvic girdle. *If the tail* is grasped, a break occurs readily; later a new tail is established. A reconstituted lizard tail is not a duplicate of the original; the spinal cord is incomplete and vertebrae and muscles are unsegmented. Regenerative activities in birds and mammals are restricted for the most part to repetitive or physiological *regeneration*. Complex organs cannot be

ple brain, a flattened body with a protrusile pharynx, a digestive system, complex excretory and reproductive organs, and a tail end When a worm is cut into a number of pieces each piece ordinarily retains its original polarity and is capable of regenerating the complex components of the body which are missing in the fragment (Fig 4) If the head is removed and the anterior part of the remaining body is split by a vertical incision, two complete heads with normal brains and eyes will be regenerated on a single body By similar methods, double tails can be formed Still more bizarre configurations can be produced, depending on the manner in which incisions of the original body are made Using various chemical agents, the polarity of an isolated portion of the mid region body can be changed so that a head will be regenerated in place of a tail

Considerable attention has been given to special reserve cells in the planarian body known as formative cells (Figs 5 and 6) Because they appear

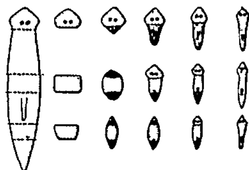


Fig 4 Regeneration in a flatworm, *Euplanaria* Portions cut from an entire worm (indicated by broken lines) gradually regenerated (dark stipple) to form entire small worms (*After Stempell from T I Storer and R L Usinger, General Zoology, 3d ed , McGraw-Hill, 1955*)

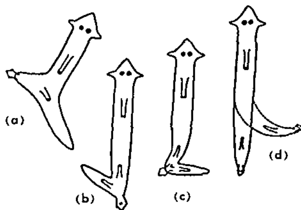


Fig 5 Organizing activity of a transplanted head in *Planaria* (*a*) A laterally implanted graft has induced a lateral outgrowth and a secondary pharynx (*b,c*) A subterminal graft has induced an outgrowth (directed forward), and two pharynges (*d*) A terminal graft has caused a reversal of polarity in the caudal part of the host, and a secondary pharynx (*After Santos from C P Raven, An Outline of Developmental Physiology, McGraw-Hill, 1954*)

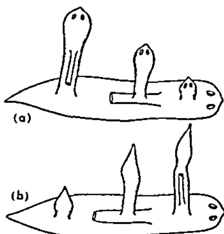


Fig 6 The effect of discontinuities in the gradient in *Planaria* (*a*) The growth caused by grafted heads is the more marked when the head is implanted more caudally (*b*) Transplanted hind ends produce strongest growth in the rostral parts of the host (*After Schewtschenko from C P Raven, An Outline of Developmental Physiology, McGraw-Hill, 1954*)

to be concerned directly with regenerative capacity, they are sometimes referred to as regeneration cells Although there is wide disagreement regarding the origin of these cells and their detailed cytological characteristics, they appear to be the primary source of the new tissues and organs formed during regeneration In different species of planaria the number of formative cells varies per unit of body volume The capacity and rate of regeneration vary directly with the frequency of these cells They have a high sensitivity to x rays Exposure of a planaria to radiation in appropriate dosage results in the destruction of the formative cells and a complete loss of regenerative capacity

Annelida. The earthworm as an example of the annelids, possesses a high degree of regenerative capacity (Fig 7) When a worm is cut into two parts a considerable number of posterior segments can be reconstituted by the anterior half Likewise, the posterior half can regenerate anterior segments, but to a more limited extent Species differences exist with respect to the number of anterior and posterior segments that can be reconstituted In some species anterior and posterior regeneration can take place simultaneously from the two ends of an isolated piece of a worm

Special reserve cells have been identified in the earthworm and are called neoblasts As in the case of the formative cells of planaria, considerable uncertainty prevails regarding their origin and the precise role they play in regeneration After removal of a portion of the body it has been observed that neoblasts migrate to the wound area and form an aggregation of cells Here they proliferate and, according to some investigators, differentiate into most of the components of the regenerated segments They do not form all of the new structures, new nerve tissue develops from proliferation of epidermis, and new intestine from the cut end of the old intestine Neoblasts are killed

by exposure to radium or x rays in appropriate dosage and the regenerative capacity of the worm is lost.

The segmental nature of the earthworm body and its high capacity for regeneration have provided a favorable field for quantitative studies of growth limitation. In one species of earthworm which has been extensively studied the number of segments which can be reconstituted posteriorly is a linear function of the distance in segments between the level of the cut and the anterior end of the animal. Regeneration ceases when the number of segments characteristic of the species has been restored.

Echinodermata. Among the echinoderms regeneration in the various types of starfishes and brittle stars is a matter of common observation to anyone who examines these animals at the seashore. All species, so far as known, readily regenerate new arms. When an animal is cut into a number of pieces a new organism develops from each piece, provided that a portion of the central disk is present. Often severe injury to an arm results in its pinching off at the base; a new arm then regenerates from the remaining short stump (*see* AUTOTOMY). In some cases a number of arms may arise from a single stump to produce a highly atypical organism. Sea urchins have a far lower regenerative capacity than starfishes but they can reconstitute various portions of the skeleton and also the tube feet. The sea cucumbers, so far as they have been studied, have been found to possess remarkable capacities for regeneration. In some species reconstitution of the body takes place after an animal has been cut into two or three pieces. Sea cucumbers exhibit an unusual phenomenon of evisceration when roughly handled. Under such circumstances they are able to regenerate a completely new set of visceral organs.

Arthropoda. *The arthropods include many divergent forms:* the various types of crustaceans, the centipedes and millipedes, the insects, scorpions and spiders. The Crustacea, especially the lobster, crab and crayfish, have been extensively studied with respect to regenerative capacity. The appendages of the Crustacea are the antennae, which are sensory in nature, the feeding appendages around the mouth, and the various types of legs used for locomotion, food procurement and protection. Also there are highly complex eyes mounted on movable stalks. All of these structures are capable of regeneration. It is often observed that the two large claws of a lobster are unequal in size. When this is the case the smaller claw is a regenerate which has replaced an original claw but has not yet grown to full size.

Considerable attention has been given to atypical regeneration in the Crustacea. For example, after amputation of an eye an antenna may be regenerated in its place. A nerve ganglion lies at the base of each optic stalk. If the ganglion is undisturbed at the time an optic stalk is amputated, a new optic stalk and eye will be regenerated. If the optic ganglion is severely injured or is removed at the time of amputation, an antenna will be regenerated instead of an eye. Other instances of atypical regeneration include the formation of a walking leg in place of one of the mouth parts, or an abdominal leg in place of a thoracic leg.

Compared with the Crustacea, regeneration in other forms of arthropods has been little studied. Some centipedes, millipedes and spiders can regenerate legs. In adult insects little regeneration occurs, although in some species legs can be reestablished. In the larval and pupal stages of insects regenerative capacity is far higher than in the adult.

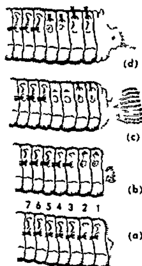


Fig. 7. Four stages (a, b, c, d) in the regeneration of the rostral end in *Sabella*. As regeneration proceeds, the foremost four segments change from the abdominal into the thoracic type by the loss of their old hooks and bristles and the formation of new ones in another position. This change proceeds in a rostro-caudal succession. (*After Berrill and Mees from C. P. Raven, An Outline of Developmental Physiology, McGraw-Hill, 1954*)

REGENERATION IN VERTEBRATES

Among vertebrates the amphibians possess the highest capacity for regeneration. Various species are able to regenerate limbs, parts of the eye, the tail, the lower jaw, and a number of other structures. Many types of fish can regenerate fins, some portions of the gills, barbules and scales. In the reptiles the principal structure which is capable of regeneration is the tail of some lizards. The tail possesses a special "breaking point" between two vertebrae posterior to the pelvic girdle. If the tail is grasped, a break occurs readily. Later a new tail is established. A reconstituted lizard tail is not a duplicate of the original; the spinal cord is incomplete and vertebrae and muscles are unsegmented. Regenerative activities in birds and mammals are restricted for the most part to repetitive or physiological regeneration. Complex organs cannot be

reconstituted, although feather and hair replacement, periodic growth of antlers in the deer, the restoration of skin, and the repair of bone fractures can all be regarded as regenerative activities in the broad sense of the term

Amphibian limb regeneration. The factors governing the regeneration of the limbs of urodeles have been subjected to extensive experimental analysis After amputation of a salamander limb, rapid coverage of the wound takes place by migration of epidermal cells from adjacent areas Within a few days an aggregation of cells is established beneath the epidermis at the tip of the limb to form a regeneration blastema From the cells of the blastema develop all of the new structures of the regenerate, except the epidermis which covers it and nerves which grow into it from the cut ends of the original nerves The establishment of a blastema is essential for regeneration Unlike the situation in some of the invertebrates, special reserve cells are not involved in its formation The regeneration blastema arises as a result of a complex series of histological changes in the tissues (connective tissues, skeleton, and muscle) at the amputation surface Injury of these tissues at the time of amputation is the primary inciting cause of blastema establishment

The regeneration blastema. A newly established regeneration blastema is an aggregation of undifferentiated cells Its transformation into a new limb resembles in many respects the normal development of a limb in an embryo (Fig 8) As its component cells proliferate, the blastema increases in size and changes in configuration from a bulb or cone into a paddle-shaped structure During growth of the blastema, its cells undergo differentiation into the new skeleton, muscle, and connective tissues which are laid down in a pattern duplicating the structures that were originally present

It has been shown by irradiation studies that the cells which make up the regeneration blastema are local in origin and do not migrate from a distance to the regeneration area A portion of a limb subjected to localized x radiation in appropriate dosage completely loses the capacity to regenerate When a limb is amputated through an irradiated

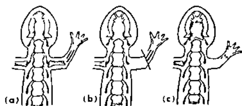


Fig 8 Diagram of forelimb regeneration in *Triton* (a) Normal skeleton of the forelimb (b) Extirpation of the humerus, followed by the amputation of lower arm and hand (c) The skeleton of the regenerated lower arm and hand is normal, but the extirpated humerus is not regenerated (*After Przibram from C P Raven An Outline of Developmental Physiology, McGraw Hill, 1954*)

region, no blastema will be established Amputation either above or below this region leads to blastema formation and normal limb regeneration

Local factors. The establishment, growth, differentiation, and morphogenesis of the blastema are governed by both local and organismic factors Local factors, especially the level of amputation determine the primary characteristics of the regenerate When a limb is amputated through the upper arm, the blastema which is established possesses the potentialities for forming all structures which are normally present below that level, a blastema which is established after amputation through the wrist possesses the potentialities for forming wrist and hand structures Interaction between the tissues of the limb stump at the level of amputation and the cells of the blastema represents the mechanism by which local factors operate *See* ANIMAL MORPHOGENESIS

Organismic factors. These factors are primarily neural and endocrine in character A limb deprived of nerves loses the capacity to regenerate If all of the nerves going to a salamander forelimb are cut in the shoulder region and the limb is then amputated, epithelial wound healing occurs, but no blastema is established Such a situation prevails as long as the limb remains nerveless Nerves themselves can regenerate, and as they grow back into a limb from the shoulder region and reach the level of amputation a blastema will be established and regeneration of the limb will ensue

Neural influence Although considerable research has been concerned with the nature of the neural influence in blastema formation, the precise mechanism of nerve action is still unknown It is unspecific in character, any type of nerve—sensory, motor, or sympathetic—can support blastema formation provided that a sufficient quantity of nerve fibers is present at the tip of an amputated limb Nerves which ordinarily supply structures other than the limb, if they are experimentally directed into a limb, will induce blastema formation Rigid quantitative requirements prevail A certain minimum number of nerve fibers must be present before regenerative activity will begin, if the number of fibers is below the minimum level no blastema will form The number of requisite fibers to support regeneration varies at different levels of a single limb

Although a sufficient quantity of nerve fibers in a limb is essential for blastema formation, it is not necessary for later phases of limb regeneration In the successive series of changes through which a blastema passes as it undergoes growth differentiation and morphogenesis it becomes emancipated from neural influence Nerve fibers are concerned particularly with the early mobilization of cells and this results in the establishment of a blastema, and the initiation of differentiation and morphogenetic activities within the blastema, neural influence is unessential for the later growth of the regenerate

Endocrine factors The endocrines most thor o ghly st d ed with respect to their influence on l mb regeneration are the pituitary the adrenal and the thyroid glands A salamander from which the p tuitary has been removed is incapable of limb regenerat on As in the case of nerves pituitary hormones are concerned primarily with the early p r egenerat on particularly the establish

A hyperthyro u s a e s

manders either by feeding thyroid extract or treat ing them with thyroxine solution Considerable controversy prevails regarding the results of such treatment. Some investigators find that limb regeneration is accelerated in such animals and others that it is retarded Much appears to depend on the type of thyroid material used and the time t is adm nistered When a hypothyroid condition

governing regeneration much additional research will be required before its precise role is deter mined See THYROID GLAND

In contrast with adult urodeles (salamanders and newts) regeneration of limbs in adult anurans (frogs and toads) does not take place It occurs, however in larval stages the details are essen t ally the same as in larval urodeles The regenera t ve capacity ceases as an anuran larva approaches metamorphosis and its loss is associated primarily w th organismic changes endocrine and neural wh ch occur at that time By various experimental me hods it is possible to restore the regenerative capac ty to the limb of an adult frog When the nerve supply of a forelimb for example is greatly augmented by surgically leading into it the great s atic nerve of the hind limb the amputated fore-

as a high regenerative capacity Although a re enerat on blastema forms after amputation of a a l and resembles in general characteristics the bla tema of an amputated limb many features of

axial organs as the spinal cord vertebral column and segmentally arranged muscles As a tail regen erat on blastema forms a new spinal cord rapidly grows into it Various experiments demonstrate the importance of the spinal cord in inducing the dif fe ent at on of skeletal and muscular components n the bla tema If a section of the spinal cord in

ing skeletal and muscular components

Amphibian ocular regeneration The eye in cer tain amphibians is capable of regenerating the lens the iris and the retina If the lens is removed from the eye a new lens will be formed by a bud like growth from the upper free margin of the iris This does not take place in all species of am phibians and has not been observed in other verte brates Lens regeneration occurs only when the normal lens is completely removed from the eye Substances given off by a normal lens inhibit the regenerative capacity of the iris removal of a lens releases the inhibition Although only a single lens ordinarily regenerates, under certain experimen tal conditions multiple lenses are formed

In some amphibians when the entire iris is re moved it will be reconstituted from the edge of the pigment epithelium of the retina When the lens and iris are simultaneously removed the iris, as it regenerates gives rise to a new lens as well In a number of species of amphibians both uro deles and anurans the entire neural retina can be regenerated from the retinal pigment epithelium

It is especially noteworthy that after regenera tion of the various ocular structures mentioned above the reconstituted eye of the amphibian func tions as well as the normal eye

Nerve regeneration The regeneration of neural structures in various types of organisms constitutes a special subject which cannot be dealt with here in detail In those invertebrates which have well organized nervous systems and which possess the ability to regenerate major portions of the body neural structures often undergo extensive recon stitution Among the vertebrates portions of the brain and spinal cord can regenerate only to a very limited degree On the contrary in many verte brates peripheral nerves are capable of extensive regeneration

Peripheral nerves in vertebrates extend to all parts of the body and end in such structures as sense organs muscles and glands The essential structural and functional unit of a nerve is the nerve fiber Each nerve is composed of large num bers of fibers bound together by connective tissues Around each individual fiber are one or more sheaths The most intimate is a noncellular myelin sheath outside of this is a thin cellular sheath the sheath of Schwann

An individual nerve fiber is a protoplasmic ex tension of a single nerve cell or neuron Although a fiber may extend a considerable distance away from the main cell body where the nucleus of the cell is located the fiber still remains an integral part of the neuron and participates in its meta bolic activities The cell bodies of neurons are situ ated in or near the main axis of an organism most of them are in the brain or spinal cord others in various types of nerve ganglia

When a nerve is cut, the fibers on the side of the cut away from their central connections undergo degeneration For example, if a nerve in the upper arm is severed, all of the fibers from the point of severance to the tips of the digits degenerate and die However, the fibers above the cut remain alive, because they retain their connections with cell bodies in the brain, spinal cord, or ganglia

Regeneration of peripheral nerves has been extensively studied in amphibians and mammals After a nerve has been severed the cut ends of its living fibers, by a type of ameboid activity, send out protoplasmic extensions or growth cones This represents the beginning of regeneration of each individual fiber Growth of protoplasmic strands from the many fibers which make up a nerve results in the regeneration of the nerve as a whole One of the primary problems in nerve regeneration concerns the manner in which fibers are directed along paths corresponding to the original pattern Many hypotheses have been proposed Chemotropism (neurotropism) does not appear to be involved The most commonly accepted hypothesis is one of contact guidance, which holds that the tips of growing nerve fibers cling to and follow various types of interfaces present in the substrate through which they grow These interfaces may be strands formed by Schwann sheath cells from preexisting fibers Scar tissue, forming from connective tissue where a nerve has been cut, often presents a serious interference to nerve regeneration [E G B]

Bibliography A E Needham *Regeneration and Wound Healing*, 1952, M Singer, The influence of the nerve in regeneration of the amphibian extremity, *Quart Rev Biol*, 27(2) 169–200, 1952, C S Thornton (ed), *Regeneration in Vertebrates*, 1959, B H Willier, P A Weiss, and V Hamburger (eds), *Analysis of Development*, 1955

Regeneration (engineering)

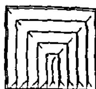
The process of feeding back a portion of the output signal of an amplifier to its input in such a way that the input signal is reinforced The result is greatly increased amplification The feedback must be positive, that is, the two signals must be in phase, and it must be limited in magnitude to prevent the circuit from going into oscillation *See* FEEDBACK CIRCUIT

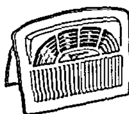
In storage devices for computers regeneration involves the restoration of deteriorating electrostatic, magnetic or other conditions to their original state This is particularly essential in charge storage cathode-ray tubes to overcome natural decay effects, as well as loss of charge by reading out the information stored *See* STORAGE TUBE

In the nuclear power field regeneration involves the purification of contaminated nuclear fuel for reuse *See* NUCLEAR FUELS REPROCESSING [J MR]

Register, air

A device attached to an air distributing duct for the purpose of discharging air into the space to be heated or cooled These openings are referred to



Wall diffusers are usually located at a low wall position to send air upward, parallel with a wall (*From Summer Air Conditioning, by S Konzo, J R Carroll, H D Bareither, The Industrial Press*)

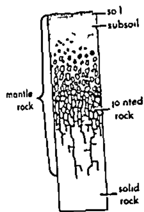
as registers, diffusers, supply outlets, or grills By common acceptance, a register is an opening provided with means for discharging the air in a confined jet, whereas a diffuser is an outlet which discharges the air in a spreading jet Both registers and diffusers may be placed at a number of locations in a room, including the floor, baseboard, low on the sidewall, window sill, high on the sidewall, or ceiling

For heating, the preferred location is in the floor, at the baseboard, or at the low sidewall of the outside wall preferably under a window For cooling the preferred location is high on the inside wall or the ceiling For year round air conditioning in homes, a compromise location is either the floor, baseboard or low sidewall at the exposed wall, especially if adequate air velocity in an upward direction is provided at the supply outlet

A well designed register effectively conceals the hole at the end of the duct, throws or projects the air in the direction and at the distance desired limits the velocity usually to 500 ft/min or slower, deflects the air away from walls and obstructions. The register also adjusts the direction of air flow to provide on the spot manipulation of the air stream, and adjusts the air flow rate to lesser amounts It should accomplish these functions without producing dust streaks on nearby walls and ceilings disturbing air noise, or large pressure losses Many registers diffusers slots and air panels are commercially available and satisfy a majority of these qualifications *See* COMFORT CONTROL, WARM AIR HEATING SYSTEM [S KO]

Regolith

The mantle rock or blanket of unconsolidated rocky debris of any thickness that overlies bedrock Undisturbed regolith may grade from agricultural soil at the surface, through fresher and coarser products of rock weathering to solid bedrock tens or even hundreds of feet beneath Elsewhere bedrock may be covered by transported soil and rock debris deposited in such forms as flood plains and deltas sand dunes beaches and bars moraines and grav



Residual regolith

foot of steep slopes and

[C.F.S.S.]

Regularia

The name given by G Cuvier in 1817 to an assemblage of echinoids in which the anus and periproct lie within the apical system The test is globular and preponderantly radially symmetrical and the ambulacral plates are commonly compound The group included in effect all those echinoids which did not fall in the Irregularia The Irregularia however have been shown to be polyphyletic by J Durham and R Melville (1957) and consequently neither they nor the Regularia constitute valid taxa. *See* ECHINOIDEA, IRREGULARIA

[H B F]

Regulation

The process of maintaining a quantity or condition essentially constant despite variations in such factors as line voltage and load In an industrial process-control system, the speed temperature, voltage, or position of a critical element can be kept constant by measuring the condition being regulated and feeding back into the system a signal representing the difference between the actual and the desired quantities For example if the temperature of a mixture of chemicals is too low a sensing element feeds back to the controller a signal that results in the application of more heat *See* CONTROL SYSTEMS

The term regulation is also used in the opposite sense, to indicate the difference between the maximum and minimum voltages at the terminals of a tube, transformer generator or other device over the range of normal operating conditions *See* VOLTAGE REGULATION [J MR]

Regulator

A control device designed to maintain the value of some quantity substantially constant Thus, a temperature regulator is a device designed to maintain the temperature of some environment at a constant value The value to be maintained can usually be established at any value within the range of the regulator by making an appropriate setting

A regulated system is a feedback control system employing a regulator to maintain some quantity of the system at a constant value. Another example is the voltage regulator system of an automobile *See* CONTROL SYSTEMS [J A H]

Reheating

The addition of heat to steam of reduced pressure after the steam has given up some of its energy by expansion through the high pressure stages of a turbine (see STEAM TURBINE) The reheater tube banks are arranged within the setting of the steam generating unit in such relation to the gas flow that the steam is restored to a high temperature. Under suitable conditions of initially high steam pressure and superheat, one or two stages of reheat can be advantageously employed to improve thermodynamic efficiency of the cycle (*See* STEAM GENERATING UNIT, SUPERHEATER, VAPOR CYCLE.) [F C E.]

Reindeer

A New World caribou *Rangifer caribou*, of the family Cervidae Reindeer are still present in limited numbers as wild animals, with two subspecies in Lapland and Greenland, and they are abundant as domesticated animals in northern Europe Siberia and Alaska In the reindeer, as in other caribou both sexes have antlers. In Alaska



The North American reindeer, *Rangifer caribou* length to 6 ft (*From E L Palmer, Fieldbook of Natural History McGraw Hill 1949*)

the reindeer is used primarily for food, in Lapland it provides the basis for an entire society, serving as food clothing and beast of burden In North America the reindeer, *Rangifer tarandus*, an introduced species has hybridized rather freely with the native caribou It feeds upon lichens, grasses, and other available plants *See* ARTIODACTYLA, CARIBOU, LICHENS [J D B]

Reinforced concrete

Portland cement concrete with steel embedded in it to assist in carrying loads Steel plays a major role in reinforced concrete structural members for several reasons It is elastic yet has considerable reserve strength beyond its elastic limit In compression it is about twenty times stronger than concrete Its tensile strength is nearly the same as its compressive strength whereas concrete is very weak in tension

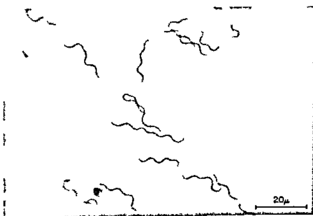
As reinforcing for concrete steel in several of the following forms may be used bars or rods wire pipe and structural shapes such as wide flange beams Bars are the most common form they generally range in size from ¼ in to 2¼ in in diameter They are specified by number No 2 being a nominal ¼ in diameter No 3 ⅜ in No 4 ½ in No 5 ⅝ in and so forth

Loads are transferred between concrete and steel by the bond along the surface of intersection The bond may be substantially improved mechanically by giving reinforcing bars raised surfaces in which case the bars are called deformed bars Deformed bars conforming to American Society for Testing Materials Specification A305 are permitted to transmit more stress through bond than are nondeformed or plain bars *See* CONCRETE CONCRETE COLUMN [F S M]

Relapsing fever

An acute infectious disease caused by various species of the genus *Borrelia* a spirochete The disease is characterized by episodes of fever which subside spontaneously and recur over a period of weeks Epidemiologically two types are recognized a louse borne type which often occurs in epidemics and a tick borne type which usually is endemic The clinical characteristics of the two are similar *See* SPIROCHETE

Type species characteristics The causative spirochete of which *B recurrentis* is the type species is highly flexible and actively motile It varies in length from 8 to 30 microns (μ) in thickness from 0 3 to 0 5 μ and has 5 10 irregular and loosely



Relapsing fever spirochetes in human blood stained with fuchsin (*General Biological Supply House Inc Chicago*)

wound spirals *Borrelia* cannot be cultured on artificial media but will grow well in the chick embryo Mice rats and guinea pigs are susceptible At refrigerator temperature and in appropriate media motility and infectivity may be maintained for months Virulence is preserved for years at approximately −76°C the temperature of dry ice A large number of species differentiated primarily on the basis of geographical distribution or the transmitting vector have been isolated and found to be the causative agents of the relapsing fevers

Relapsing fever in man Following an incubation period of about 7 days the initial attack starts abruptly with chills high fever headache and often pains in the muscles and joints It lasts 2–8 days and ends by crisis A remission period of 3–10 days is followed in untreated cases by a relapse similar to the initial attack but milder There may be 3–10 relapses Mortality in the endemic infection varies between 2 and 5% but in epidemics it may reach 50% *See* EPIDEMIOLOGY

Serum agglutinins and bactericidal antibodies are demonstrable in both the experimental and human disease Definitive diagnosis is made by demonstration of the organisms in the blood by darkfield microscopy by examination of stained blood films or by animal or chick embryo inoculation The treatment of choice is chlortetracycline penicillin oxytetracycline and streptomycin have therapeutic value *See* CHLORTETRACYCLINE OXYTETRACYCLINE PENICILLIN STREPTOMYCIN

Transmission and prevention Transmission is solely by insect vectors In northern and western Africa Europe and parts of Asia the disease is spread mainly by the body louse *Pediculus humanus* Infection occurs when lice are crushed near a bite or scratch that provides a portal of entry for the organisms In the endemic areas of Central and South Africa Asia and the Americas the disease is commonly tick borne the most important vector being the genus *Ornithodoros* many species of which have been shown to be infected in nature Transovarian infection which is transmission of the infective agent from the female to the embryo through the egg occurs in the tick

There is no effective vaccine nor are there practicable chemoprophylactic measures In louse-borne epidemics the isolation and large scale treatment of patients and mass delousing of the population with DDT are effective procedures DDT resistant lice have appeared in some countries In endemic areas reduction of the tick population by periodic spraying of living quarters with benzene hexachloride may effect a major reduction in the incidence of the disease For taxonomy *see* SPIROCHAETALES [T B T]

Relative motion

All motion is relative to some frame of reference The simplest laboratory frame of reference is three mutually perpendicular axes at rest with respect to an observer Such a system is commonly used in the laboratory when various types of motion are

being studied The general effects of other motions to which the system as a whole is subjected are then neglected and the system is said to be isolated In terms of the frame of reference of an observer some distance from Earth, the laboratory frame of reference would be moving with Earth as it rotates on its axis and as it revolves about the Sun What would be a simple form of motion in the laboratory frame of reference would appear to be a much more complicated motion in the frame of refer ence of the distant observer *See* FRAME OF REFER ENCE

Motion means continuous change of position of an object with respect to an observer To another observer in a different frame of reference the object may not be moving at all or it may be moving in an entirely different manner The motions of the planets were found in ancient times to appear quite complicated in the laboratory frame of ref erence of an observer on Earth By transferring to

terms of elliptical orbits The validity of one description is no greater than the other, but the latter description is far more convenient *See* PLANET

Relative velocity. That motion is relative to an observer must have been implicit in the earliest ideas of motion In the mechanics of Galileo and Isaac Newton these ideas became clarified and methods were developed for finding the relative velocity of two bodies each moving with a differ ent velocity If the velocity of one body is v_1 rep re ented in Fig. 1 by the magnitude and direction of the vector v_1 and if a second body has a veloc ity v_2 represented by the vector v_2 then v_3 is the vector to be added to v_1 to make the sum equal to v_2 and consequently the vector v_3 is equal to the vector difference $v_2 - v_1$ Therefore the relative velocity of the second body with respect to the first is that velocity represented in magnitude and di rection by the vector v_3 all vectors being drawn to a suitable scale

In the simplest case the velocities v_1 and v_2 are parallel (Fig 2) The relative velocity of the sec ond body with respect to the first is again v_3 and its magnitude is the difference in the numerical values of v_2 and v_1 When the two velocities are

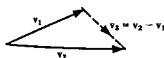


Fig 1 Relative velocity of two bodies moving at an angle with respect to one another

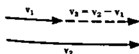


Fig 2. Relative velocity of two bodies moving in the same direction

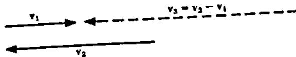


Fig 3 Relative velocity of two bodies moving in opposite directions

anti parallel (in opposite directions, Fig 3), one is negative with respect to the other and the magnitude of the relative velocity of v_2 with respect to v_1 is again the vector difference v_3, which is now the numerical sum of the two To obtain the relative velocity of v_1 with respect to v_2, the arrow head on v_3 should be reversed For example, when two automobiles move with velocities of 40 and 60 mph respectively in the frame of reference of an observer stationed by the roadside, the relative velocity of the two is 20 mph If they were moving in opposite directions, the relative velocity would be 100 mph

The flight of an airplane illustrates the princi ples of relative motion The common statement that the speed of an airplane is, for instance, 300 mph is essentially a meaningless statement since the speed is not designated with any particular frame of reference in mind and the listener must make an assumption that it is perhaps the ground speed in still air For the airplane to sustain itself with nor mal lift it is the air speed or speed with respect to the air that is important In a head wind of 100 mph the airplane would have a ground speed of 200 mph In a similar tail wind it would have a ground speed of 400 mph If the wind velocity is at an angle to the direction of flight, the relative velocities must be considered The velocity with respect to the ground would be the vector sum of the velocity of the plane with respect to the air and the velocity of the wind with respect to the ground *See* FLIGHT CHARACTERISTICS

Relative acceleration Acceleration, like velocity is relative to the observer's frame of reference An automobile starting from rest is accelerated with respect to the Earth but the driver of the car does not see himself or the car accelerated forward He sees objects at rest with respect to the roadway accelerated backward with respect to himself and the car Persons by the roadside see him and the car accelerated forward In the driver's frame of reference the car is at rest

Acceleration is a vector quantity involving both magnitude and direction just as velocity is Just as the velocities of two objects may be represented by vectors and their relative velocity obtained by subtracting one vector from the other so also the accelerations of two bodies may be represented by vectors and the relative acceleration of one with respect to the other obtained by taking the vector difference

Relativity. Since according to Einstein's theory of relativity the velocity of light is the limiting velocity that any physical object can attain, the ad dition or subtraction of very large velocities cannot

be accomplished by Galilean Newtonian methods, and the rules to be followed are derived from relativistic theory *See* RELATIVITY [R D RU]

Relativistic electrodynamics

The study of the interaction between charged particles and electric and magnetic fields when the velocities of the particles approach that of light Relativistic electrodynamics, which can be considered as an extension of the everyday laws of electricity and magnetism, is an important consideration in high energy particle accelerators, in high current, high voltage vacuum tubes, and in electromagnetic radiation

The laws which relate the electric and magnetic fields to the charges and currents which produce them are known as Maxwell's equations Charged particles or current elements in such fields experience a force which is called a Lorentz force The motion of these charged particles and current elements in such fields is then determined by Newton's laws, appropriately generalized for relativistic velocities *See* ELECTRON MOTION IN VACUUM, MAXWELL'S EQUATIONS, RELATIVISTIC MECHANICS, RELATIVITY

J C Maxwell's contribution to relativistic electrodynamics was to formulate his four equations and to introduce the concept of displacement current (*see* DISPLACEMENT CURRENT) Although each equation was originally deduced for static or other restrictive conditions Maxwell implied the validity of each for fields which varied in arbitrary ways with time

The Lorentz force describing the influence of the electric field **E** and magnetic field **B** on a moving charge q of velocity **v** is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$$

or
$$F_x = q(E_x + v_y B_z - v_z B_y), \text{ etc}$$

Together with the equation of continuity (*see* EQUATION OF CONTINUITY)

$$\operatorname{div} \mathbf{J} + \frac{\partial \rho}{\partial t} = \operatorname{div} (\rho \mathbf{v}) + \frac{\partial \rho}{\partial t} = 0$$

or
$$\frac{\partial J_x}{\partial x} + \frac{\partial J_y}{\partial y} + \frac{\partial J_z}{\partial z} + \frac{\partial \rho}{\partial t} = 0$$

and Newton's second law

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d}{dt}\left(\frac{m_0 \mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}\right)$$

or
$$F_x = \frac{dp_x}{dt} = \frac{d}{dt}\left(\frac{m_0 v_x}{\sqrt{1 - \frac{v^2}{c^2}}}\right), \text{ etc}$$

one has a self consistent set of equations to determine the fields and motions of charges under various conditions In the preceding equations, ρ is the electric charge density, **J** is the current density, **p** is the momentum of the charge, m_0 its rest mass, and c the velocity of light

One direct consequence of Maxwell's equations comes from the fact that the fields **E** and **B**, in the absence of charges and currents, satisfy a wave equation with wave velocity

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} \approx 3 \times 10^8 \text{ m/sec}$$

The fact that this is the same as the velocity of light led Maxwell to infer that light is an electromagnetic wave phenomenon, thus joining the fields of electricity, magnetism, and optics

Invariance of Maxwell equations. The property of the Maxwell equations that makes them applicable to problems in relativistic electrodynamics is their relativistic invariance Specifically this means that Maxwell's equations will seem correct to an observer traveling with a constant velocity as well as to an observer at rest One must realize however, that a magnetic field in the rest frame will appear to be both an electric and a magnetic field in the moving frame *See* FRAME OF REFERENCE

As an example, consider a charge moving through an externally applied static electric and magnetic field with a velocity such that the Lorentz force vanishes In this case

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}) = 0 \qquad \mathbf{E} = -\mathbf{v} \times \mathbf{B}$$

In a frame of reference moving with velocity **v**, the charge will seem to be at rest and will therefore experience a force $\mathbf{F}' = q\mathbf{E}'$, responding only to the electric field in the new system Since the electron now remains at rest, the force **F'**, and therefore **E'**, must vanish (A magnetic field **B'** will still be present) It is therefore apparent that the electric and magnetic fields must change when a Lorentz transformation is made (*see* LORENTZ TRANSFORMATIONS) The fields mix in such a way as to make Maxwell's equations appear to be the same to all observers

Relativistic beams. One interesting consequence of the way in which the fields transform is the diminution of the repulsive force between charges moving with high velocity in parallel paths In the reference system moving in the z direction with velocity v, in which the charges are at rest, the magnetic field vanishes and the electric field of one at the location of the other is

$$E'_y = \frac{q}{y'^2}$$

where y' is the separation of the charges In the rest system, the fields can be shown to transform to

$$E_y = \frac{E'_y}{\sqrt{1 - \frac{v^2}{c^2}}} \qquad B_x = -\frac{v}{c^2}\frac{E'_y}{\sqrt{1 - \frac{v^2}{c^2}}} \qquad y' = y$$

so that the force between the charges is

$$F_y = q(E_y + vB_x) = qE_y\left(1 - \frac{v^2}{c^2}\right) = qE'_y\sqrt{1 - \frac{v^2}{c^2}}$$

The force is thus reduced by a factor $1-(v^2/c^2)$ compared with the electric force alone and by a factor $[1-(v/c^2)]^{1/2}$ compared with the electric force which exists when the charges are at rest

Another way of illustrating this cancellation is to consider the force on the outer charges in a cylindrical beam of current I consisting of charges moving with velocity v in the z direction The tangential magnetic field at the surface $(r = a)$ is by Amperes law

$$B_\theta = \frac{2\mu_0 I}{a}$$

where μ_0 is the magnetic inductive capacity of free space $\approx 1.257 \times 10^{-6}$ henry/m Continuity of charge requires the electric charge density per unit length to be

$$\tau = \frac{I}{v}$$

in which case the radial electric field at the surface is, by Gauss law

$$E_r = \frac{2I}{\epsilon_0 v a}$$

where ϵ_0 is the electric inductive capacity of free space $\approx 8.85 \times 10^{-12}$ farad/m The force on a charged particle at the surface is therefore

$$F_r = q(E_r - vB_\theta) = \frac{2Iq}{\epsilon_0 a v}\left(1-\frac{v^2}{c^2}\right)$$

nal or z motion) of this charge configuration is determined from

$$\frac{d^2r}{dt^2} = \frac{F_r}{m_t} = \frac{F_r}{m_0}\left(1-\frac{v^2}{c^2}\right)^{1/2} = \frac{2Iq}{\epsilon_0 m_0 a v}\left(1-\frac{v^2}{c^2}\right)^{3/2}$$

where
$$m_t = m_0\left(1-\frac{v^2}{c^2}\right)^{-1/2}$$

The from

... the charges were at rest

An important application of this effect is to relativistic beams of particles which might be expected to disperse as a result of space charge

the last equation Another way of looking at the phenomenon is that in the rest system of the electrons, the accelerator appears extremely short because of the Lorentz contraction and the space-charge forces have little time to spread the beam apart. (The argument is actually more complicated since one must deal with an accelerated frame of reference, but the conclusion is the same)

The tendency for relativistically charged beams not to diverge is of course not restricted to accelerators It is important in consideration of electron optics in high current, high voltage vacuum tubes, and may even be important in controlled fusion devices Applications have even been considered where the transverse motion of ion beams is made convergent by introducing an opposing beam of electrons Partial charge neutralization results in a decrease of the electric defocusing force but leaves the magnetic focusing force of almost the same magnitude unchanged thus providing a considerable net focusing action In addition it is often possible to diminish the space-charge defocusing of positive ion beams by space charge neutralization using electrons

Other relativistic phenomena As a result primarily of relativistic dynamics many of the usual formulas for interaction of particles and fields are altered Some of these altered expressions are listed in the accompanying table

The significance of the phenomena included in

ple to the usual

these phenomena are encountered There are many other applications in which the relativistic aspects are not small corrections but are indeed the main considerations involved Included in this group are most of the particle accelerators with energies in the relativistic range (for example, electrons with energy above 1 Mev protons with energy above 1 Bev)

Linear particle accelerators Most electron linear accelerators produce extremely relativistic electrons (more than 500 Mev in the traveling wave accelerator at Stanford University compared with the rest energy of ½ Mev) The considerations regarding the transverse motion of the electrons have already been described the magnetic interaction between the moving electrons reduces the repulsive force by several orders of magnitude and no other means of controlling the transverse motion is necessary The longitudinal motion is also inhibited by the large mass of the electrons Indeed the phase oscillations which are responsible for longitudinal stability become extremely slow, since all electrons are effectively traveling at the same velocity c they neither fall behind nor overtake one another and are therefore less inclined to get out of phase In fact the phase motions are inversely proportional to the longitudinal mass so called because the acceleration is in this case in the direction of the initial velocity

$$m_l = \frac{m_0}{\left(1-\dfrac{v^2}{c^2}\right)^{3/2}}$$

which is extremely large in the relativistic range

Similar consideration applies to relativistic linear ion accelerators (above 1 Bev) although these of necessity have a large section in which nonrela

Relativistic form of common equations

Effect	Nonrelativistic form	Relativistic form
1 Acceleration of charge through potential difference V	$E_{kin} = \frac{m_0 v^2}{2} = qV$	$E_{kin} = m_0 c^2 \left[\frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} - 1 \right] = qV$
	$v = \sqrt{\frac{2E_{kin}}{m_0}} = \sqrt{\frac{2qV}{m_0}}$	$v = \sqrt{\frac{2E_{kin}\left(1 + \frac{E_{kin}}{2m_0c^2}\right)}{m_0\left(1 + \frac{E_{kin}}{m_0c^2}\right)}} = \sqrt{\frac{2qV}{m_0}} \sqrt{\frac{1 + \frac{qV}{2m_0c^2}}{1 + \frac{qV}{m_0c^2}}}$
	$p = m_0 v = \sqrt{2m_0 E_{kin}}$ $= \sqrt{2m_0 qV}$	$p = \frac{m_0 v}{\sqrt{1 - \frac{v^2}{c^2}}} = \sqrt{2m_0 E_{kin}\left(1 + \frac{E_{kin}}{2m_0c^2}\right)}$ $= \sqrt{2m_0 qV} \sqrt{1 + \frac{qV}{2m_0c^2}}$
2 Circular motion in a uniform magnetic field	$\frac{m_0 v^2}{r} = qvB$	$\frac{m_0 v^2}{r\sqrt{1 - \frac{v^2}{c^2}}} = qvB$
	$p = m_0 v = qBr$	$p = \frac{m_0 v}{\sqrt{1 - \frac{v^2}{c^2}}} = qBr$
	$\omega = \frac{v}{r} = \frac{qB}{m_0}$	$\omega = \frac{v}{r} = \frac{qB}{m_0}\sqrt{1 - \frac{v^2}{c^2}} = \frac{qBc}{\sqrt{q^2B^2r^2 + m_0^2c^2}}$
	$E_{kin} = \frac{q^2B^2r^2}{2m_0}$	$E_{kin} = \sqrt{q^2B^2r^2c^2 + m_0^2c^4} - m_0c^2$
3 Langmuir Child law	$J = \frac{4\epsilon_0}{9}\sqrt{\frac{2q}{m_0}} \frac{V^{3/2}}{L^2}$	$J = \frac{4\epsilon_0}{9}\sqrt{\frac{2q}{m_0}} \frac{V^{3/2}}{L^2}\left(1 - \frac{3}{28}\frac{qV}{m_0c^2}\right)$

tivistic phenomena predominate since conventional ion sources yield particles in the Mev range

Circular particle accelerators The cyclotron is basically a nonrelativistic device with a uniform field in which a fixed radio frequency accelerates particles according to the relation

$$\omega = \frac{qB}{m} = \frac{qB}{m_0}\sqrt{1 - \frac{v^2}{c^2}}$$

As long as m is nonrelativistic ($\cong m_0$) the rotational fre quency $\omega_{rot}/2\pi$ of the particles is the same as the applied radio frequency and they are accel erated Once the particles become relativistic how ever ω_{rot} and ω_{rf} are no longer the same to accel erate the particles beyond this region either the frequency must be made to decrease with time (synchrocyclotron) or the fields must be made to increase with radius (one of the features of the fixed field alternating gradient or FFAG cyclo tron)

The synchrotron (both electron and proton) is a device which has a magnetic field varying in time in such a way as to keep the orbit radius constant Specifically the equations

$$B = \frac{mv}{qr} = \frac{m_0\omega}{q}\left(1 - \frac{\omega^2 r^2}{c^2}\right)^{-1/2} \qquad \omega = \frac{v}{r}$$

give the necessary relation between the magnetic field and the frequency for a given radius In the case of relativistic electrons $v \approx c$ and only the magnetic field need be varied to accelerate the par ticle

A betatron is an electron accelerator in which orbits of constant radius are maintained without radio frequency fields The electrons are acceler ated by an increase of magnetic flux linking the or bit the relation between the linked flux and the field at the orbit may be derived from Faraday s law of induction and from the relativistic relation for circular motion Faraday s law is

$$\text{emf} = -\frac{d\Phi}{dt}$$

indicating that the particle gains a kinetic energy

$$\Delta E_{kin} = q\frac{\Delta\Phi}{\Delta t} = \frac{qv}{2\pi r}\Delta\Phi$$

during each turn The symbol Δ refers to the change per turn and Φ is the flux linked by the orbit The particle will remain in the same orbit provided the magnetic field and momentum in crease in such a way that

$$\Delta p = qr\,\Delta B$$

One obtains

$$\Delta E_{kin} = F\,\Delta x = \frac{\Delta p}{\Delta t}\Delta x = \frac{\Delta x}{\Delta t}\Delta p = v\,\Delta p$$

This condition between the increase of energy and momentum per turn can therefore be met only if

$$\frac{\Delta\Phi}{\pi r^2} = 2\,\Delta B$$

that is if the change in the average field enclosed by the orbit is twice the change of the field at the orbit This equation is known as the betatron con dition For additional information on the betatron

condition and detailed information on particle motion in all the important types of particle accelerators, see PARTICLE ACCELERATOR

Electromagnetic radiation Another important area of application for relativistic phenomena is in the field of electromagnetic radiation In fact the phenomenon of radiation itself involves a complete relativistic treatment, as one may safely assume from the appearance of c, the velocity of light, in all relevant formulas

When one considers a configuration of current sources varying sinusoidally in time, the low frequency solution corresponds to the coherent superposition of the radiation fields from each element of exciting current As the frequency is increased an interesting relativistic aspect of the problem appears disturbances from different current elements take different times to reach the point of observation, and the relative phase of these contributions to the total signal is changed One can take this retardation into account by adding to the phase an amount $kr_{e,o}$ where $r_{e,o}$ is the distance between the element of current under consideration and the observation point. The wave number k is given in terms of the frequency by

$$k = 2\pi f/c = 2\pi/\lambda$$

where λ is the wavelength of the radiation Clearly, this effect will be important if the phase difference between opposite extremes of the current distribution is comparable with or greater than 2π, that is, if $kD \geqq 2\pi$ or $D \geqq \lambda$ Here D is a typical linear dimension of the current distribution In general retardation effects must be taken into account if the dimensions of the radiating system are comparable with or greater than the wavelength

Synchrotron radiation As a further example of a purely relativistic phenomenon, consider the radiation of a charged particle which undergoes acceleration during circular motion at constant speed. The classical Larmor formula for slowly moving charges gives for the rate of energy radiation per second

$$R = \frac{2}{3}\frac{q^2}{4\pi\epsilon_0}\frac{a^2}{c^3}$$

where a is the particle acceleration The appropriate relativistic generalization in the case of circular motion is

$$R = \frac{2}{3}\frac{q^2}{4\pi\epsilon_0}\frac{a^2}{c^3}\left(1 - \frac{v^2}{c^2}\right)^{-4}$$

where $a = v^2/r$, r being the orbit radius One can see directly the radical change for relativistic velocities. This radiation called synchrotron radiation, represents a serious limitation on the magnitude of energies attainable in circular electron accelerators.

[R L C]

Bibliography W W Harman, *Fundamentals of Electronic Motion*, 1953, M S Livingston, *High Energy Accelerators* 1954, L. Page and N I Adams, Jr., *Principles of Electricity*, 3d ed 1958, W K H Panofsky and M Phillips, *Classical Electricity and Magnetism*, 1955, J C. Slater and N H Frank, *Electromagnetism*, 1947, A Sommerfeld, *Electrodynamics*, Vol 3, 1952, J A Stratton, *Electromagnetic Theory*, 1941

Relativistic mechanics

Any form of mechanical theory compatible with either the special or the general theory of relativity The term is usually interpreted in a narrower sense to refer to the mechanics of a system of particles moving in a given external field, or of a perfect fluid, subject to the special theory The present discussion is limited to these cases

The existence of a class of equivalent inertial reference systems is admitted in the special theory of relativity as in Newtonian mechanics The space and time variables of any two inertial frames are related by a Lorentz transformation (see LORENTZ TRANSFORMATIONS) In the equations given in this discussion cgs units are used for mechanical quantities and Gaussian units for the electromagnetic field with c designating the speed of light in free space

First law of motion. According to Albert Einstein a particle subject to no forces moves in a straight line with constant speed, and a particle which moves with the speed of light in one inertial reference frame does so in all equivalent systems.

A free particle of rest mass m_0, moving with velocity $\mathbf{v}$ has momentum $\mathbf{p}$ and total energy E, with

$$\mathbf{p} = \frac{m_0\mathbf{v}}{\sqrt{1 - v^2/c^2}} \qquad E = \frac{m_0c^2}{\sqrt{1 - v^2/c^2}} = c\sqrt{(m_0c)^2 + \mathbf{p}^2}$$

The internal energy of the particle is $E_0 = m_0c^2$, and its kinetic energy is $T = E - E_0$

A particle of zero rest mass ($m_0 = 0$), such as a photon, can move only with the speed of light, its energy momentum relation being $E = pc$ Conversely a material particle with nonvanishing rest mass cannot attain a speed equal to that of light. The concept of kinetic energy is not applicable to particles of zero rest mass A photon moves with the speed of light and has a frequency ν such that its energy is $E = h\nu$ where h is Planck's constant

The preceding relations can be extended to a system of particles which undergo elastic or inelastic collisions with each other provided the particles move independently of each other between collisions The laws of conservation of momentum and energy are

$$\Sigma\mathbf{p} = \text{constant} \qquad \Sigma E_i = \text{constant}$$

It is not necessary that the total number of particles remain constant, so that particles may be created or annihilated in a collision

If the symbols $\{m'_i\}$ indicate the rest masses of the particles entering into a collision and $\{m''_j\}$ the rest masses of those emerging from the collision the energy release in the collision is

$$Q = (\Sigma_i m'_i - \Sigma_j m''_j)c^2$$

The numerical value of Q may be positive, negative, or zero

Second law of motion. The Einstein Planck law states that the equation of motion of a particle of charge ϵ in a given external electromagnetic field $(\mathbf{E}, \mathbf{H})$ is

$$\frac{d\mathbf{p}}{dt} = \epsilon\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{H}\right)$$

The rate at which the field does work on the particle is given by

$$\epsilon\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{H}\right) \cdot \mathbf{v} = \epsilon\mathbf{E} \cdot \mathbf{v}$$

The energy equation of the particle is $dT/dt = \epsilon\mathbf{E} \cdot \mathbf{v}$

Longitudinal and transverse mass Sometimes it is found convenient to express the Einstein Planck equation in a form similar to Newton's equation of motion of a particle The acceleration vector, $\mathbf{a} = d\mathbf{v}/dt$, is written in the form $\mathbf{a} = \mathbf{a}_\parallel + \mathbf{a}_\perp$, where $\mathbf{a}_\parallel$ is parallel to the velocity vector, while $\mathbf{a}_\perp$ is perpendicular to it The force vector is written similarly as $\mathbf{F} = \mathbf{F}_\parallel + \mathbf{F}_\perp$, where $\mathbf{F}_\parallel$ and $\mathbf{F}_\perp$ are parallel and perpendicular to the velocity vector, respectively The Einstein Planck equation of motion is equivalent to the following pair of equations

$$\frac{m_0}{[1 - (v/c)^2]^{3/2}}\mathbf{a}_\parallel = \mathbf{F}_\parallel, \quad \frac{m_0}{[1 - (v/c)^2]^{1/2}}\mathbf{a}_\perp = \mathbf{F}_\perp$$

The coefficients

$$m_l = \frac{m_0}{[1 - (v/c)^2]^{3/2}}, \quad m_t = \frac{m_0}{[1 - (v/c)^2]^{1/2}}$$

are called the "longitudinal mass" and the "transverse mass" of the particle respectively

Uniformly accelerated motion As an example of motion according to this law, consider a particle which starts from rest at $t = 0$ and is accelerated in a homogeneous electrostatic field $\mathbf{E}$ Taking the x axis along the direction of the field with the origin at the initial position of the particle, the equation of motion is

$$\frac{d}{dt}\left[\frac{m_0 v}{\sqrt{1 - v^2/c^2}}\right] = \epsilon|\mathbf{E}|$$

with $v = dx/dt$ The solution of this equation can be written in the form

$$\left(\frac{\epsilon|\mathbf{E}|x}{m_0 c^2} + 1\right)^2 - \left(\frac{\epsilon|\mathbf{E}|t}{m_0 c}\right)^2 = 1$$

Owing to the mathematical form of this solution the motion is frequently referred to as being hyperbolic For small values of the variable $(\epsilon|\mathbf{E}|t/m_0 c)$ this formula agrees with that obtained from Newton's second law of motion, namely $x = (\epsilon|\mathbf{E}|/2m_0)t^2$ With increasing time, the speed of the particle approaches the speed of light asymptotically, as the inertia of the particle increases without bound

Four dimensional formulation The relativistic second law of motion of a particle, and its equation of energy, can be written in a unified form by use of space time coordinate notation The coordinates will be written as $x^0 = ct$, $x^1 = x$, $x^2 = y$, $x^3 = z$, this notation being abbreviated to $\{x^\alpha\} = (ct, \mathbf{r})$ Every 4-vector can be expressed, in a fixed inertial frame, as a combination of a time component and a 3-vector

The line-element in space time

$$(ds)^2 = (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$$

is invariant in numerical value and in algebraic form under Lorentz transformations When the space variables are identified with the coordinates of a moving particle, it takes the reduced form $ds = c\sqrt{1 - v^2/c^2}\,dt$

The 4-velocity is defined by the relations $u^\alpha = dx^\alpha/ds$, so that

$$\{u^\alpha\} = \left(\frac{1}{\sqrt{1 - v^2/c^2}}, \frac{\mathbf{v}/c}{\sqrt{1 - v^2/c^2}}\right)$$

The 4-momentum of the particle is $\{p^\alpha\} = \{m_0 c u^\alpha\}$

The 4-dimensional form of the law of motion is

$$dp^\alpha/ds = f^\alpha \qquad (\alpha = 0,1,2,3)$$

where $\{f^\alpha\}$ is the 4-vector representing the force acting on the particle The time component $(\alpha = 0)$ of this set of equations is the energy equation, while the space components $(\alpha = 1,2,3)$ are the equations of motion

On comparison of these expressions with the Einstein Planck law of motion it is found that for a charged particle in an external electromagnetic field

$$\{f^\alpha\} = \left(\frac{\epsilon\mathbf{E} \cdot \mathbf{v}}{c^2\sqrt{1 - v^2/c^2}}, \frac{\epsilon\left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{H}\right)}{c\sqrt{1 - v^2/c^2}}\right)$$

The fact that this expression transforms like a 4-vector under Lorentz transformations follows from the properties of the field vectors This provides a proof of the form invariance of the Einstein Planck law under Lorentz transformations

The 4-dimensional form of the law of motion can be considered to be a generalization of the Einstein Planck law if the force 4-vector is assigned arbitrarily

Relativistic hydrodynamics. The basic postulate that no mechanical or electromagnetic influence can be propagated in space with a speed greater than that of light precludes the introduction of the concept of a rigid body into the special theory of relativity Every continuous body must have infinitely many degrees of freedom For this reason hydrodynamic theory is the basic model for treating continuous media Important differences from the corresponding Newtonian theory of hydrodynamics arise from the necessity of taking into account the kinematical requirements of relativity theory The relativistic theory of hydrodynamics has not

of relativity and to the study of cosmological models of the universe

The behavior of a perfect fluid can be characterized in terms of three physical quantities (1) mass density ρ, (2) momentum density g and (3) the internal stress system (p_{ij}) (see FLUID-FLOW PROPERTIES) These variables are connected by four scalar differential equations, namely, the equation of continuity and the equations of motion The equations are formulated in such a manner that they resemble the corresponding equations of Newtonian hydrodynamics as closely as possible, and are form invariant under Lorentz transformations

It will suffice here to give the 4-dimensional form of the equations of relativistic hydrodynamics. The energy momentum density of the fluid is represented by a symmetric tensor of second rank $[T^{\alpha\beta}]$ which is defined in terms of the physical variables of the fluid by the matrix equation

$$[T^{\alpha\beta}] = \begin{bmatrix} c^2\rho & cg_x & cg_y & cg_z \\ cg_x & p_{xx} & p_{xy} & p_{xz} \\ cg_y & p_{yx} & p_{yy} & p_{yz} \\ cg_z & p_{zx} & p_{zy} & p_{zz} \end{bmatrix}$$

Under a Lorentz transformation the components of this tensor transform according to the formula

$$T'^{\alpha\beta} = \sum_{\lambda=0}^{3}\sum_{\mu=0}^{3} \frac{\partial x'^{\alpha}}{\partial x^{\lambda}} \frac{\partial x'^{\beta}}{\partial x^{\mu}} T^{\lambda\mu}$$

... motion of the fluid in the absence of external force fields, are expressed by the four differential equations

$$\sum_{\beta=0}^{3} \frac{\partial T^{\alpha\beta}}{\partial x^{\beta}} = 0 \qquad (\alpha = 0,1,2,3)$$

See RELATIVISTIC ELECTRODYNAMICS RELATIVITY, SPACE TIME

[E L HI]

Bibliography A Einstein, *The Meaning of Relativity* 1956 G Joos *Theoretical Physics* 3d ed 1958.

Relativity

A theory of the physical meaning of space and time due largely to Albert Einstein In its present form, the theory consists of two parts (1) the special ...

one another ... at constant velocity relative ... and (2) the general theory (1916)

... physics.

SPECIAL (RESTRICTED) THEORY

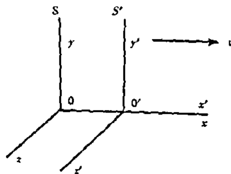
The principle of *relativity* as formulated in 1899 by H Poincare, and extended by him in 1904, stated the impossibility of determining the absolute motion of a physical body, or of a physical reference system, by dynamical, optical or electromagnetic means This principle was accepted by Einstein as the starting point of the special theory of relativity, and was associated with a fundamental interpretation of the meaning of space and time as determined by physical measurement

The existence of a class of uniformly moving (*inertial*) *reference systems* is admitted in the special theory of relativity, as in Newtonian mechanics and the theory is stated only for these systems It is convenient, for simplicity of expression, to associate with each inertial frame of reference an observer who may be thought of as making observations on physical phenomena in terms of the coordinates of his frame The space and time variables of inertial observers are connected by a group of mathematical transformations known as the Lorentz transformations. See LORENTZ TRANSFORMATIONS

The principle of relativity is interpreted in a mathematical sense to imply that the equations of physics must be form invariant under the transformations of the Lorentz group. The basic physical assumption is that *no mechanical or electromagnetic influence can transport energy from point to point in space with a speed exceeding that of light, approximately* 3×10^{10} cm/sec

Moving frames of reference. The nature of the theory can be examined by considering the special case of two observers using similarly oriented reference frames which move relatively along a common (x,x') axis (see illustration) System S', with space and time coordinates (x',y',z',t'), moves to the right with speed v with respect to *system* S *in which the coordinates are* (x,y,z,t) The observer associated with each frame interprets his time variable in the usual sense as measuring time simultaneously at all points in his reference system

If the coordinates are adjusted so that the origins of the two frames are in coincidence at the instants $t = 0$, $t' = 0$, as noted by the observers in



Moving reference systems

S and S', the transformation equations connecting the space and time variables of the two systems are

$$x' = \frac{x - vt}{\sqrt{1-(v^2/c^2)}} \qquad y' = y \qquad z' = z$$

$$t' = \frac{t - vx/c^2}{\sqrt{1-(v^2/c^2)}}$$

$$x = \frac{x' + vt'}{\sqrt{1-(v^2/c^2)}} \qquad y = y' \qquad z = z'$$

$$t = \frac{t' + vx'/c^2}{\sqrt{1-(v^2/c^2)}}$$

Here c is the speed of light These equations reduce to the corresponding Galilean transformation equations of Newtonian mechanics, $x' = x - vt$, $y' = y$, $z' = z$, $t' = t$ in the limit $v/c \to 0$ It is an important epistemological feature of Einstein's theory that relations of this type can be established by consideration of physical measurements on moving bodies by means of light signals

Two relations of particular interest will be considered as illustrative of the meaning of space and time variables as formulated by Einstein (1) the contraction of a body in the direction of its motion (FitzGerald Lorentz contraction), and (2) the slowing of the rate of a moving clock (time dilatation)

FitzGerald Lorentz contraction. Consider a physical body *which is at rest in system S', and which* has an exterior surface F' which is described by the equation $\varphi(x',y'z') = 0$ To an observer in S, this body appears to be transported to the right with speed v To determine the apparent form of the surface F of the body as observed in S, the primed variables are eliminated from the equation of F' *The equation of F is found to be*

$$\psi_t(x,y,z) = \varphi\left(\frac{x - vt}{\sqrt{1-(v^2/c^2)}}, y, z\right) = 0$$

For example, a sphere of radius a in S', with equation $(x')^2 + (y')^2 + (z')^2 - a^2 = 0$ appears to an observer in S to be the moving ellipsoid

$$\left(\frac{x - vt}{a\sqrt{1-(v^2/c^2)}}\right)^2 + \left(\frac{y}{a}\right)^2 + \left(\frac{z}{a}\right)^2 = 1$$

with semiaxes $(a\sqrt{1-(v^2/c^2)}\ a,a)$ It follows that the surface F can be obtained from F' by a uniform contraction in the direction of motion by a factor $\sqrt{1-(v^2/c^2)}$

The famous Michelson Morley experiment can be interpreted, in the sense of Einstein's ideas as showing (among other things) the physical reality of the contraction of the interferometer in the direction of its motion For a detailed discussion of the Michelson Morley *experiment*, *see* LIGHT

Rate of a moving clock (time dilatation). Consider two events which occur at a fixed point in S', but at different times, t'_1 and t'_2, with $t'_2 > t'_1$ As observed from S, these events take place at different points in space (x_1,y,z) and $(x_2\,y,z)$, and at different times t_1 and t_2, such that $x_2 - x_1 = v(t_2 - t_1)$ It is found from the transformation equations connecting S and S' that

$$t_2 - t_1 = \frac{t'_2 - t'_1}{\sqrt{1-(v^2/c^2)}}$$

Since $t_2 - t_1$ and $t'_2 - t'_1$ have the same algebraic sign, the order of the two events in time is the same in S' and S However, the time interval between the events appears to the observer in S to be longer than it does to the observer in S' This result is interpreted by the statement that a moving clock appears to run at a slower rate than an identical clock at rest, in the ratio $\sqrt{1-(v^2/c^2)}$, where v is the speed of the clock *See* CLOCK PARADOX

A modification of the Michelson Morley experiment, known as the Kennedy Thorndike experiment, has been developed for the study of the time-dilatation effect The results of the experiment were in complete agreement with the theory An indirect verification of the effect has been provided by the observation that the measured lifetime of high energy mesons increases with energy

Four-dimensional formulation. The mathematical development of the special theory is much facilitated by a 4-dimensional formulation which was initiated by H Poincare and stated more completely by H Minkowski (*see* SPACE TIME) Let the variables $x^0 = ct$, $x^1 = x$, $x^2 = y$, $x^3 = z$ be *considered as coordinates in a 4-dimensional space* time In the following equations every Greek index takes the values 0, 1, 2, 3, while a repeated index indicates a summation over these values

Under a Lorentz transformation the differentials $\{dx^\alpha\}$ transform according to equations $dx'^\alpha = L_\beta{}^\alpha\, dx^\beta$, such that the quantity

$$(ds)^2 = \eta_{\alpha\beta}\, dx^\alpha\, dx^\beta$$
$$= (dx^0)^2 - (dx^1)^2 - (dx^2)^2 - (dx^3)^2$$

is invariant both in numerical value and in algebraic form Here the quantities $(\eta_{\alpha\beta})$ are defined by the relations

$$\eta_{00} = +1 \qquad \eta_{11} = \eta_{22} = \eta_{33} = -1$$
$$\eta_{\alpha\beta} = 0 \quad \text{if} \quad \alpha \neq \beta$$

A repeated Greek index indicates a summation over all values of the index from 0 to 3

If $d\mathbf{r} = dx^1\mathbf{i} + dx^2\mathbf{j} + dx^3\mathbf{k}$ is the infinitesimal displacement vector connecting neighboring points on the trajectory of a particle and if the particle velocity is $\mathbf{v} = d\mathbf{r}/dt$, then $ds/c = \sqrt{1-(v/c)^2}dt$ In an inertial frame in which the particle is instantaneously at rest $(\mathbf{v} = 0)$ ds/c is equal to the time interval dt measured by the observer in that frame

The motion of a material particle between space points $(x_0\ y_0\ z_0)$ and $(x_1\ y_1\ z_1)$ which it occupies at times t_0 and t_1 in a fixed inertial system, can be described in space time coordinates by a set of equations $x^\alpha = x^\alpha(\theta)$ where the parameter θ $(0 \leqq \theta \leqq 1)$ is adjusted so that $\{x^\alpha(0)\} = (ct_0, x_0\ y_0, z_0)$ and $\{x^\alpha(1)\} = (ct_1, x_1\ y_1, z_1)$ This 4-dimensional

curve is called the world line of the particle between the given points in space-time The integral

$$\int \frac{ds}{c} = \int_0^1 \frac{ds}{d\theta}\frac{d\theta}{c}$$

is well defined and is referred to as the lapse of proper-time of the particle between the given points. It is usually assumed to be equal to the time interval which would be measured by a physical clock accompanying the particle

For a particle moving with the speed of light

ticles of zero rest mass, such as photons The world line of a particle of nonvanishing rest mass is not singular at any point *See* RELATIVISTIC MECHANICS

The 4-velocity of a material particle is defined by the relations $u^\alpha = dx^\alpha/ds$, and its 4-acceleration $a^\alpha = du^\alpha/ds$ These quantities obey the conditions

$$\eta_{\alpha\beta}u^\alpha u^\beta = (u^0)^2 - (u^1)^2 - (u^2)^2 - (u^3)^2 = 1$$
$$\eta_{\alpha\beta}a^\alpha u^\beta = a^0u^0 - a^1u^1 - a^2u^2 - a^3u^3 = 0$$

These definitions provide a means of transforming kinematical quantities under Lorentz transformations.

The formulation of the equations of mechanics, electrodynamics, thermodynamics and hydrodynamics in 4-dimensional form reveals internal symmetries in their structures which are not otherwise apparent. Even when it is desirable, for clarity of

a fundamental physical entity was of great importance in Einstein's extension of the special to the general theory of relativity From this starting point he developed his basic idea that the gravitational field is representable in terms of an intrinsic geometrical structure of space-time [E L.HI]

GENERAL THEORY

This is the relativistic theory of gravitation, published by Einstein in 1916 It superseded the classical theory of gravitation which had been formulated by Isaac Newton almost 300 years earlier Newton's theory had been outstandingly successful in explaining the motions of all celestial bodies under the influence of their mutual gravitational attraction, but based as it was on the prerelativistic concepts of space and time Newton's theory was inconsistent with the new concepts of space and time associated with the restricted theory of relativity

The general theory of relativity rests on the so called principle of equivalence which asserts the equivalence of gravitational forces and inertial forces, that is, the forces experienced by accelerated observers, such as persons in a rolling ship, for example, or in an automobile coming to a sudden stop Capitalizing on the principle of equivalence, the general theory of relativity modifies further the space and time concepts of special relativity and arrives at the notion of the curved space time continuum The flat space time of special relativity remains a special case in which the curvature is zero and it represents physically the absence of gravitation

Once the geometric properties of space time were fused with the physical phenomenon of gravitation, the general theory *of relativity proceeded to* the formulation of a new law for the gravitational field Einstein's law of gravitation, which was supplemented by a new law of motion for bodies subject to *gravitational forces*

These new concepts, all of which were proposed by Einstein in 1916, have been the subject of further intensive exploration ever since Because of the complexity of the laws, the search for mathematical solutions, that is fields that obey the laws, is a continuing challenge. *Also the predictions of* the new theory to the extent that they differ from those of Newton's classical theory have become the subject of numerous experimental investigations

Principle of equivalence. This is based on the observation that in free fall, all bodies undergo the same acceleration (*see* FREE FALL) Newton explained this fact in terms of his law of gravitation In this law the mass plays a dual role on the one hand the mass is the ratio between the force *acting on a body and the resulting acceleration*, *on* the other the gravitational force itself is proportional to the mass of the body that is falling In these two roles mass is frequently referred to as inertial mass and as gravitational mass, respectively Newton assumed that for all material bodies these two masses are equal The first confirming experiment of high accuracy was performed in 1890 by R Eotvos his findings were confirmed and extended by several other workers, such as L Southerns in 1910, and P Zeeman in 1917 *See* GRAVITATION

If it is true that in a gravitational field all bodies *are accelerated* at the same rate (which is approximately 980 6 cm/sec² on the surface of the earth), then gravitation shares this peculiarity with so called inertial fields According to classical (prerelativistic) physics inertial fields *are not true* force fields at all, that is to say, they do not represent interactions between material bodies. An inertial field arises whenever the observer is himself in accelerated motion Newton's first law of motion, to the effect that in the absence of external forces bodies remain in their states of rest or of uniform rectilinear motion is valid only if their motions are *described with respect to an unaccelerated observer an inertial observer Otherwise* the peculiarities in the motion of the observer are reflected in the (apparent) observed accelerations of objects

not subject to external forces Examples of such external forces are centrifugal forces Coriolis forces and the effects observed inside vehicles that accelerate or decelerate suddenly

Einstein's elevator Because both in free fall and in inertial fields the observed accelerations of material objects are independent of their masses or other individual properties it is difficult if not im possible to tell them apart by means of experi ments or observations based on local conditions To elucidate this point Einstein considered a win dowless elevator whose cables had been severed and which was falling freely in its shaft If air re sistance could be neglected then the elevator itself would fall at the same rate as a passenger or other objects in its interior Thus if the passenger were to release a ball from his hand the ball would not approach the floor but would remain floating in mid air because the rate of its fall would equal that of the elevator In this respect then the in terior of the freely falling elevator would resemble conditions in interstellar space

Conversely if the elevator were brought into in terstellar space and if an imaginary being were to pull with constant force at a cable attached to the roof then conditions in the interior would resem ble those experienced in a field of gravity at rest In this case an inertial field would masquerade as a gravitational field

The principle of equivalence asserts that the ob servable local effects of inertial and gravitational fields are indistinguishable that to this extent they are equivalent

The principle of equivalence does not assert equivalence of inertial and gravitational fields in all respects For instance if one were to make a survey of the magnitude and direction of the force field in an extended domain then one could deduce from the data obtained the presence or absence of a gravitational field In the absence of a gravita tional field it is possible to prescribe the motion of an observer in such a manner that for him the ap parent accelerations of bodies not subject to other forces (for example electric or magnetic) vanish throughout the domain surveyed in the presence of a gravitational field this is impossible In other words the gravitational field according to the principle of equivalence is characterized by its in homogeneities

Einstein s law of gravitation In geometrical language if one cannot construct a frame of refer ence (that is to say an observer equipped with measuring rods clocks and other facilities to per form kinematic experiments) in which gravita tional inertial fields vanish throughout an extended domain then that space time continuum is curved Whereas the curvature of a 2 dimensional manifold (an ordinary surface) is characterized by a single parameter at each point a 4-dimensional manifold such as the space time continuum requires 20 pa rameters at each point Only when all 20 vanish is the continuum flat Einstein conjectured that these 20 parameters contained the intrinsic characteris tics of the gravitational field those that would not depend on the state of motion of the observer and hence that the laws of the gravitational field would have to place some restrictions on these 20 param eters of curvature There remained the task of find ing such restrictions that the resulting physical laws would reduce to those of Newton s theory in all situations in which Newton s law was known to yield the correct answers Einstein succeeded in showing that the desired laws consisted of the re quirement that 10 of the 20 parameters should van ish in the absence of other physical fields or matter and that in their presence these 10 parameters should be proportional to the density of mass (or energy) to the density of linear momentum and to the stress in various directions (which includes the pressure) respectively These relationships which take the same mathematical form for all observers together are known as Einstein s law of gravitation

Principle of covariance In classical physics as well as in the special theory of relativity the laws of physics are formulated so as to hold with respect to a certain class of observers or frames of refer ence the so called inertial observers or inertial frames of reference These are the observers with respect to whom force free bodies are unacceler ated According to the principle of equivalence it is impossible to identify an inertial frame of refer ence by locally conducted experiments at least in the presence or suspected presence of a gravita tional field Alternatively one would have to admit all conceivable observers or frames of reference regardless of their state of motion for a descrip tion of nature and the formulation of its laws Fre quently therefore the asserted equivalence of all frames of reference is called the principle of equiv alence and this formulation of the principle is just as valid as the one presented previously

The mathematical equivalent of this physical principle is the requirement that the laws of nature should take the same mathematical form in all con ceivable curvilinear coordinate systems if arbi trary new coordinates are substituted for an origi nal set of coordinates the differential equations representing all the laws of nature are to reproduce themselves automatically This formal requirement is known as the principle of covariance Einstein s law of gravitation satisfies this principle Most physicists agree that any future developments or modifications of Einstein s theory will have to sat isfy the principle of covariance as well

Law of geodesic motion The classical Newto nian theory of gravitation consists of two distinct sets of laws One law which is specific for gravita tion states that the attractive force between two bodies is proportional to the r masses and inversely proportional to the square of the distance between them the other law (Newton s second law of mo tion) states that the resulting acceleration of each body is proportional to the force acting on it and inversely proportional to its mass Einstein s law of gravitation corresponds to the first of these two laws The second the law of motion takes a form

that corresponds to the (local) equivalence of gravitational and inertial forces in the general theory of relativity in the absence of nongravita

ifold (so-called geodesic lines) in the presence of other fields, such as electromagnetic fields, these are to determine the curvature of a trajectory For relatively slow motions this requirement results in a motion that is very close to the motion corresponding to Newton's second law The law of geodesic motion also obeys the principle of covariance

Later developments in the theory of motion The original principle of geodesic motion leaves something to be desired in that any body of finite mass by its very presence contributes to the curvature of space time, and in a manner dependent on its motion Hence it would appear as if trajectories and local geometry depend on each other in a manner that prevents the determination of either In 1937 Einstein L. Infeld and B Hoffmann succeeded in showing that the geodesic principle is unnecessary for the determination of a particle's trajectory in general relativity It is sufficient to require that outside the particle itself, in empty space Einstein's law of the gravitational field be satisfied The particle is then constrained to move in the usual fashion without the separate formulation of

would hold in a wide class of conceivable theories satisfying the requirements of the principle of covariance.

mental tests requires some knowledge of the solutions Efforts have gone in two directions One approach has been to search for special but exact solutions the other to construct approximate but very general solutions The exact solutions are considered first.

Schwarzschild's solution K Schwarzschild in

[illegible] sphere At considerable distances from the center of the mass point or sphere, Schwarzschild's solution is indistinguishable from the Newtonian gravitational potential which is inversely proportional to the distance from the center At close distances (the so called Schwarzschild radius or gravitational rad[illegible]) the

[illegible] gravitational [illegible] [ap]proaches the speed of light Because actual ponderable bodies in nature are not mass points but are extended in space, conditions at the Schwarzschild radius are not observable If the entire mass of the earth were concentrated in one point, then its Schwarzschild radius would be 1 cm

Neither Newton's nor Einstein's theory of gravitation forbids the occurrence of negative masses Such masses would repel, rather than attract, other bodies of positive mass Negative masses have not as yet been observed in nature Unless they are discovered at some future date, the failure of all known theories of gravitation to exclude them must be considered a defect.

Other static solutions H Weyl and T Levi Civita during the years 1917–1921 discovered solutions corresponding to fields that might be produced by masses arranged along an axis of symmetry in arbitrary fashion Their solutions are significant chiefly because outside of Schwarzschild's solution they are the only static solutions as yet discovered (except for approximate solutions) They include a so-called dipole field, one that would be produced by a positive and a negative mass of equal magnitudes in close proximity to each other No physical field of this type has ever been observed though electric and magnetic dipoles are well known

Gravitational waves In 1937, Einstein and N Rosen discovered solutions that could be interpreted as waves produced by oscillating ponderable matter along an infinitely long cylindrical axis Such waves would spread from their source with the speed of light, they could be observed, in

could never occur in nature the Einstein Rosen waves are of considerable theoretical interest as no one had as of early 1960 succeeded in obtaining rigorous gravitational waves spreading from a central region of small extension in all directions Between 1958 and 1960 H Bondi, F Pirani, and I Robinson and independently N Rosen and A Peres, published rigorous solutions that represent plane and plane fronted waves These investigations are continuing Several investigators, foremost among them J Weber are looking into the possibility of discovering gravitational waves experimentally Their discovery and perhaps their production in the laboratory, would not only be a welcome confirmation of one of the predictions of the theory, but would be of great significance for the whole of physics

Approximate solutions Because rigorous solutions are available only in very special cases approximate solutions have been discussed widely, this despite the fact that in many cases the existence and the properties, of an approximate solution do not guarantee the existence of a similar rigorous solution With the help of approximate solutions most of the work on the theory of motion has been done waves of more general types than

those of Einstein and Rosen have been constructed and examined, and the properties of possible sources have been classified (P Bergmann and R Sachs, 1958)

Observational tests of the theory. By and large, the general theory of relativity makes but a few predictions that deviate from those of Newton's theory of gravitation Its principal advance is that it places the theory of gravitation on foundations that are compatible with the remainder of modern physics (*see* QUANTUM FIELD THEORY, UNIFIED FIELD THEORIES) The principal predictions of general relativity that permit experimental verification are (1) a modification in the orbit of Mercury, (2) the deflection of light rays passing close to the sun, and (3) a slight decrease in the frequency of spectral lines of radiating physical systems located in places of large gravitational potential

Orbit of Mercury Because of the slight perturbations of one planet on another, all planetary orbits rotate very slightly, so that the locations of their apsides change in the course of time In the case of Mercury general relativity predicts an additional rotation of the apsides of about 43 sec of arc per century which in spite of its smallness, has been verified with satisfactory accuracy

Deflection of light According to general relativity light rays passing close to the limb of the sun suffer a slight deflection which at best amounts to less than 2 sec of arc Such a deflection resulting in the apparent displacement of stars from their normal locations in the sky, can be observed only during total eclipses of the sun Since 1919 every total eclipse has been exploited for this purpose with the result that in spite of the difficulties of observation, the predicted effect has been verified with an accuracy approaching 5%

Gravitational red shift A shift of spectral lines toward the red has been observed both on the sun where the effect is masked by many other disturbances and also in the case of a few very massive and dense stars the so called white dwarfs In 1956 F Singer suggested an experiment involving the installation of an atomic clock aboard an artificial satellite J Zacharias suggested somewhat earlier a purely terrestrial experiment involving the comparison of two atomic clocks at different altitudes Finally early in 1960, J Schiffer, T Cranshaw, and A Whitehead in England reported the successful observation of the gravitational red shift in an experiment in which gamma rays emitted by the artificially radioactive iron isotope Fe^{57} are absorbed by a crystal containing the same isotope (Mossbauer effect) By separating emitter and absorber a few meters vertically, they demonstrated loss of absorptivity (because of slight loss of resonance), which could be restored by a very small relative velocity of just the right magnitude This method, which had also been suggested by R V Pound, may turn out to be the most accurate experimental determination of the gravitational red shift possible By comparison, the astronomical observations, though qualitatively in the right direction, are probably quantitatively unsatisfactory *See* ATOMIC CLOCK, EINSTEIN SHIFT [P G B]

Bibliography P G Bergmann, *Introduction to the Theory of Relativity*, 1942, A Einstein *On the Special and the General Theory of Relativity—for the General Reader* (1917), English tr, *Relativity, the Special and the General Theory*, 1920, A Einstein, *The Meaning of Relativity*, 5th ed, 1956, A Einstein and L Infeld, *The Evolution of Physics*, 1938, H A Lorentz, A Einstein, et al, *The Principle of Relativity*, reprint, 1952, C Møller, *The Theory of Relativity*, 1952

Relaxation oscillator

An electronic circuit which has two stable states, resulting in two distinct output levels, and which switches between the two states at a rate determined by the rate of rise or decay of voltage across the storage element in an *RC* or *RL* circuit The output waveform is usually nonsinusoidal, and may be approximately a square wave, a saw tooth wave, or a series of short repeating pulses *See* WAVE SHAPING CIRCUITS

One of the most widely used forms of relaxation oscillator is the astable multivibrator (*see* MULTIVIBRATOR) which generates a rectangular or square wave *In this circuit two devices are connected* so that they are alternately on and off Connected together by a positive feedback path they are driven rapidly from one state to the other

There is also a class of circuits in which a single device has two stable conditions, either on off or on with two distinct states or levels Switching between the two states usually involves an *RC* time constant The blocking oscillator (*see* BLOCKING OSCILLATOR) is representative of such a circuit wh [illegible]

[illegible]

or triode can function as a simple relaxation oscillator For example the circuit using the glow discharge tube shown in Fig 1 will generate a fixed amplitude periodic saw tooth waveform When the rising exponential of voltage reaches the breakdown potential V_{max} of the tube the capacitance C discharges through the tube and the voltage is lowered until the extinction potential is reached at

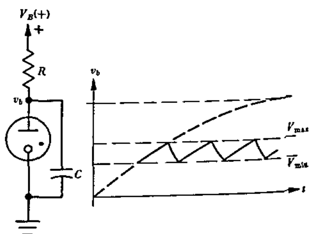


Fig 1 Glow tube relaxation oscillator

which time the cycle is repeated The gas thyratron (Fig 2) functions in the same manner except that the breakdown voltage is controllable by the grid voltage as shown by the breakdown charac- [illegible] opportunity for

[illegible]

are bistable in nature, that is, there can be two distinct voltages for a given current, or possibly two distinct currents for a given voltage, dependent upon the voltage of other control elements An external circuit containing storage elements, such as capacitance, can be made to cause the device to switch between the two conditions.

One such circuit is the historic, but little used, Van der Pol oscillator shown in Fig 3 If the suppressor grid of the pentode is highly negative, no plate current can flow, all the space current flows to the screen grid, and the screen voltage is low because of the series resistance in the screen circuit. This is one stable, or equilibrium, condition On the other hand, if the suppressor voltage is zero

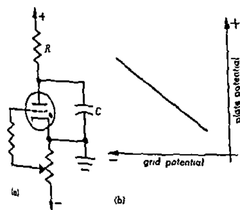


Fig 2. *Thyratron relaxation oscillator* (a) Typical circuit. (b) Breakdown *characteristic*

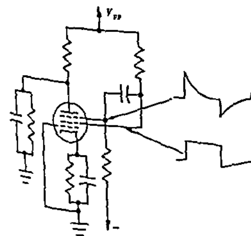


Fig 3 Van der Pol relaxation oscillator

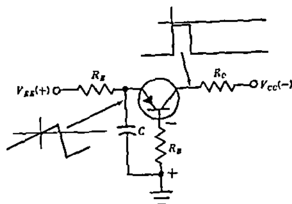


Fig 4 Point-contact transistor relaxation oscillator.

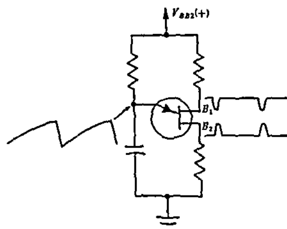


Fig 5 Unijunction transistor as relaxation oscillator

or slightly positive, most of the space current will go to the plate and the screen voltage will be higher Because of the *RC* time constant in the suppressor screen circuit, neither of these conditions is maintained, and switching action takes place between the two at a rate depending upon the rise and fall of voltage across *C*, which is a function of the *RC* time constant

Use of solid state devices. Multivibrators employing transistors find widespread use There are other relaxation oscillators using single transistors or transistorlike devices with a current gain greater than unity, such as the point contact transistor which can be arranged as a relaxation oscillator as shown in Fig 4 During the ON period a large collector current flows through R_B and R_C, making the base slightly negative, and the capacitance *C* charges toward a negative potential through the emitter circuit It finally reaches a point at which the current gain drops to less than unity The voltage drop across R_B suddenly decreases, and the emitter circuit becomes reverse biased The emitter will remain reverse biased until *C* can again charge to the conduction level through R_E The circuit is suitable for the generation of short pulses such as those needed for trigger circuits

As shown in Fig 5, a unijunction transistor, sometimes called a double base diode, may be used as a simple relaxation oscillator trigger generator [G M G]

Bibliography S Seely, *Electronic Engineering*, 1956, R F Shea, *Transistor Circuit Engineering*, 1957

Relaxation time (electrons)

The characteristic time interval required for an assembly of electrons in a metal or semiconductor to approach the equilibrium distribution after the sudden removal of a perturbation

The equilibrium distribution for such an assembly of electrons, called an electron gas, is the Fermi Dirac distribution, $f_0(\epsilon)$, where ϵ is the electron energy Under the influence of certain perturbations such as an electric field applied to a metal, the electron distribution reaches a steady state value which differs from $f_0(\epsilon)$ If this perturbation is suddenly removed, the electron gas will gradually approach a distribution given by $f_0(\epsilon)$ The processes which tend to restore equilibrium are various collision mechanisms, such as electron electron collisions, collisions of electrons with stationary imperfections, and interactions between electrons and lattice vibrations A relaxation time τ can be defined, provided the approach to equilibrium is exponential that is, provided

$$\left(\frac{\partial f(\mathbf{k},\mathbf{r})}{\partial t}\right)_{\text{coll}} = \frac{-f - f_0}{\tau}$$

Here $f(\mathbf{k}\,\mathbf{r})$ is the nonequilibrium distribution which is generally a function of the wave vector **k** of the electron and of the position **r** The wave vector **k** rather than the momentum **p** is used here because in crystalline solids the quantum mechanical state of an electron can be specified by means of **k** Frequently the relation $\hbar\mathbf{k} = \mathbf{p}$ holds where $\hbar$ is Planck's constant h divided by 2π, but this is not generally true See BAND THEORY OF SOLIDS

The conductivity of a perfect crystalline solid in which there are conduction electrons should be infinite because in such a lattice there would be no relaxation mechanisms and a current once established, would never decay The observed conductivities of metals show that the electron relaxation times at room temperature are quite short, of the order of 10^{-14} sec Even at low temperatures, the relaxation times in nonsuperconducting metals are still about 10^{-9} sec even in very pure samples These short relaxation times must be a consequence of imperfections in the crystal and the temperature dependence of the conductivity of metals indicates that at least one of the relaxation mechanisms is associated with the thermal excitation of the lattice Thermal, or lattice, scattering, arises because the thermal vibrations of the crystal's ions about their equilibrium positions destroy the perfect periodicity of the ideal lattice The perturbation which is introduced into the crystalline potential gives rise to scattering from one state **k** to another state **k'**, and the probability for such an event increases with increasing thermal agitation *See* LATTICE VIBRATIONS, RESISTIVITY, ELECTRICAL [F J B]

Bibliography S Fluegge (ed), *Handbuch der Physik*, vol 19, 1956, F Seitz and D Turnbull (eds), *Solid State Physics*, vol 4, 1957

Relay

A device, usually electromagnetic, that uses a variation of the current in one circuit to control conditions in another circuit Relays are commonly used for remote switching and control, protection of electrical devices and systems, and in digital computers All relays are essentially composed of three elements an actuating element, a movable element, and a set of contacts Relays may be classified according to time action, mechanical details or principle of operation, and application

Classification by time action. Time action refers to the length of time from the instant that the actuating element is energized to the instant that the relay contacts are closed If the time action is specified as instantaneous, the contacts are closed immediately after the current in the actuating coil exceeds its minimum calibrated value If the relay has a definite time limit, there is a definite time elapsed between the instant that the current in the actuating coil exceeds its minimum calibrated value and the instant that the relay contacts are operated This time setting should be independent of the amount of current through the actuating coil, being the same for all values of current in excess of the minimum calibrated value of the relay The time action may also be specified as inverse-time. The time delay is inversely proportional to the amount of excitation, that is, the greater the excitation the less is the time delay of relay action Practically all inverse-time relays are provided with a definite minimum time feature so that the relay never becomes instantaneous in its action

Classification by mechanical design. The most important relay designs are (1) armature or clapper type, (2) plunger (3) induction disk (4) induction cylinder, (5) beam type, (6) balanced

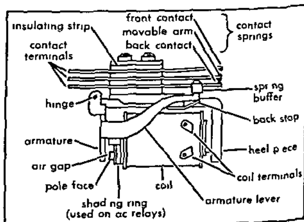


Fig 1 Armature or clapper relay construction (From K Henney and C Walsh, *Electronic Components Handbook*, McGraw Hill, 1957)

beam, (7) inductor loop (8) multiple pole induction cup (9) polar element, (10) thermal, and (11) *various* electronic types The more common types of design are discussed in the following paragraphs.

Armature or clapper type Figure 1 shows this common type of relay The electromagnetic coil attracts the pivoted armature to the pole face The armature lever, which is fastened to the armature, also moves and operates the movable contact arm If current is removed from the coil, the armature is returned to its original position by a spring The relay is shown in its unenergized position *The movable contact is making contact with the back contact but not the front contact The back contact is therefore called the normally closed* (NC) contact and the front contact the normally open (NO) contact. The number, arrangement, and sequence of operation can be varied according to the design of this or other types of relays Figure 2 shows the

field created by the solenoid, or coil, causes the plunger to rise and operate the relay contacts The relay may be designed for a variety of contact arrangements shown in Fig 2 This type of relay is especially adaptable to instantaneous action

Induction disk relay The induction disk relay Fig 4, operates on the same principle as the common watt hour meter It operates on alternating current only The disk corresponds to a short circuited secondary like the squirrel cage rotor of an induction motor The iron core is generally designed with three poles one located below the disk and two smaller poles located above the disk Various arrangements of exciting coils are possible depending upon the function to be performed In Fig. 4 *the* upper poles are excited by series current coils while the lower pole is excited by a po-

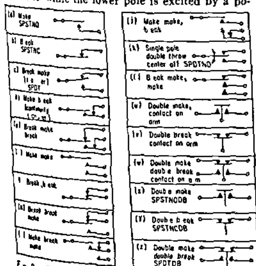


Fig 2 Contact arrangements and nomenclature of the National Association of Relay Manufacturers (From W D Cockrell, *Industrial Electronics Handbook* McGraw Hill 1958)

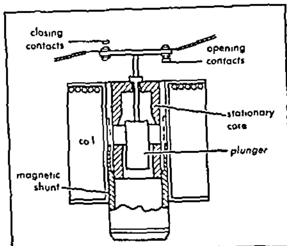


Fig 3 A plunger-type relay

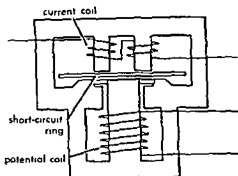


Fig 4 Induction-disk power relay

tential thus making the relay a power device Eddy currents are induced in the disk, and a resultant torque is proportional to the magnetic flux and the eddy currents The disk drives a moving contact toward a stationary contact which forms the switch for relaying purposes

An induction disk relay has an inverse time characteristic the disk rotates slowly for small values of excitation more rapidly for larger excitations A permanent horseshoe magnet provides damping *which increases with an increase in the speed of* the disk The damping therefore places an upper limit to the speed of the disk

Induction cylinder relay A typical induction cylinder relay is illustrated in Fig 5 The iron core is composed of four poles with a cylinder or cup of conductive metal suspended in the air gap Torque is developed in the rotor (the cylinder in this case) if two magnetic fields are produced at different points and if these magnetic fields are out of time *phase with each other Cylinder* rotation is used to position a movable contact in relation to fixed contacts Two diametrical poles are often considered as a single pair and excited by the same *current*, as shown by the two vertical poles These may be current or voltage excited The two other poles in the horizontal axis are shown with independent windings the excitation of which is dependent upon the particular function to be performed

Balanced beam relays A beam type of relay is illustrated in Fig 6 The armature is a beam pivoted at its center Two magnetic poles, one at each end of the balanced beam, act on the armature One pole is wound with a current winding and the other pole has two potential coils If the magnetic forces of the current and potential coils are equal, the beam will rest in its reference position If one of the poles delivers a greater magnetic pull than the other, the beam will deflect and a contactor operation will take place This relay is particularly adaptable to transmission line protection, in which the relay essentially acts to measure the line impedance from the relay position to the point of fault

Thermal relays These relays have a resistance heating element which heats a bimetallic strip One of the metals in this strip has a higher coefficient of expansion than the other Therefore, as the strip heats, it bends and makes contact with a stationary contact, as in Fig 7 The thermal relay is often used as a time delay relay It has the disadvantage of requiring a cooling period before being usable again Thermal relays are usually glass enclosed to reduce ambient temperature effects

Classification by function. Relays perform many specific functions Circuit protection relays may protect against undervalues or overvalues of current, voltage power, or frequency Protection may also be provided by the differential principle Relays are also used as interlock safety devices, stepping switches sequence controls, and time-delay devices *See* ELECTRIC PROTECTIVE DEVICES

Relay torque is generally dependent upon the reactions produced by one or more magnetic fields. To make a relay respond to any of the above circuit conditions such as current overload, the torque producing magnetic fields must be proportional to the particular quantity being monitored

Differential principle. One of the most common methods of protection of machines, equipment, and transmission lines is based upon the proposition that the input and output currents of a device should be the same so long as there is no internal trouble This is illustrated in Fig 8 for a three phase generator For simplicity the method of protection is shown applied to only one phase Two current transformers are used in each phase winding, the secondaries being connected in a closed loop with the relay operating coil as shown As long as the generator windings are in perfect condition, there will be zero current through the relay operating coil An internal machine fault will unbalance the machine currents at the two ends of each phase winding resulting in unbalanced currents in the current transformer secondaries The unbalanced current will flow through the operating coil of the relay causing it to operate Operation of the contacts will trip out the machine circuit breakers, and set in motion corrective measures [J C T]

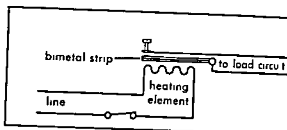


Fig 7 Thermal time-delay relay (From D M Ci sidine, Process Instruments and Controls Handbo McGraw Hill, 1957)

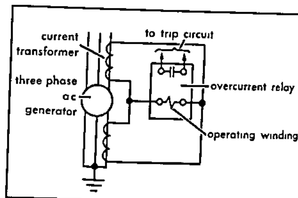


Fig 8 Differential overcurrent relay protection

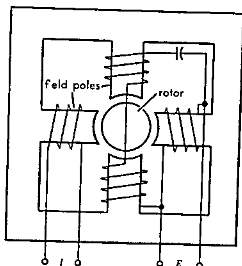


Fig 5 Induction cylinder relay

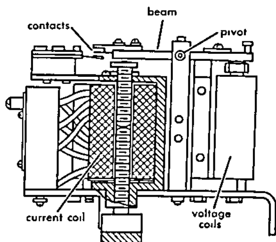


Fig 6 Beam high speed impedance relay

Bibliography K Henney and C Walsh, *Electronic Components Handbook*, 1957, J G Tarboux, *Electric Power Equipment*, 3d ed 1946

Reliability of equipment

The probability that a component part equipment or system will satisfactorily perform its intended *function under given* circumstances such as environmental conditions limitations as to operating *time and frequency and* thoroughness of maintenance.

Reliability is *influenced by all aspects of an* engineering effort the ultimate reliability of a component or a system depends upon the *quality of research* involved in its conception its design the manner in which it is manufactured the *external* influences on its operation maintenance considerations, and other factors *See* SYSTEMS ENGINEERING

The prodigious increase in the complexity of engineering systems in the last two decades emphasizes the difficult problem of achieving high reliability For example comparing dollar costs a pre-World War II civil airplane had about $4000 worth of electronic control navigation and communication apparatus The postwar commercial DC-6 required in excess of $50 000 worth of electronic apparatus, a more than tenfold increase A contemporary jet bomber has over $1 000 000 worth of electronic *gear a twentyfold increase over the* DC-6 and over 200 times that of the pre World War II airplane

Mathematical basis of reliability A quantitative *definition of reliability* should be simple enough for mathematical calculations long enough to include *significant* effects should permit the utilization of available experimental reliability data and must *lead to* results which are meaningful in terms of decisions influencing research design specifications, *manufacture* operation maintenance and to be total inoperation or must mathematically define the meaning of satisfactory performance for all aspects of the operation of the equipment

Thus in specific circumstances reliability can mean a number of different things Reliability can be considered as a function of time, the time that a piece of equipment operates satisfactorily compared with the total time over which *it was expected* to operate For example an early warning radar requires periodic *maintenance and repair* but to be adequately reliable it must be operable a large part of the total time *of attempted operation*. A special case of this definition is where the expected time of operation *coincides with the total* required time of operation that is the completion of the mission without a failure

A second specific definition of reliability is the average number of hours of *maintenance for each* hour of satisfactory performance A third definition is the fraction of the total *number of identical* equipments which operate satisfactorily over a given period of time A fourth measure of reliability compares *the actual performance of a device with* its ideal performance Actual and ideal performance may be defined for example as the performance of a device under the guidance of an average operator compared to the performance of a device under the operation of a trained and skilled operator or the field operation of a device compared with its laboratory performance This fourth definition of reliability permits a direct comparison between *the performance of devices* under idealized conditions and under actual conditions thus introducing *the influence of environment on performance* and it also considers the realization of a performance ratio *rather than a good bad or operate-nonoperate* evaluation

Mathematical theory of probability The mathematical theory of reliability is based on the statis-

1

probability of A is given by the ratio of the number of times A occurs over the number of times $A + B$ occurs *See* PROBABILITY

On the basis of this mathematical definition of probability and the results of a number of experiments the statistical reliability data for a given *component equipment or system can be presented* in several ways If N is the number of units at the *start of the test F the number of failures* at any given time t the number of survivors S at any given time is equal to N minus F and the reliability R at any given time in the test is given by S divided by N The failure rate Y is the change of reliability with time (the slope of the reliability time curve) Mathematically

$$Y = \frac{dR}{dt} = \frac{1}{N}\frac{dF}{dt}$$

The hazard rate Z is the ratio of the number of *failures per hour to the number of survivors at* that time given mathematically by

$$Z = \frac{1}{S}\frac{dF}{dt} = \frac{1}{R}\frac{dR}{dt}$$

Thus by integration the relationship between the reliability and the hazard is

$$R = \exp\left(-\int_0^t Z\,dt\right)$$

The relation between reliability failure rate and hazard is $Y = RZ$ Fig 1 shows this relation *graphically*

Joint probability Thus far the theory of probability *has been applied* to a single unit In a system the aggregate of a number of units joint probability *relates* independent failures of components to the over all reliability of the system The ability to predict joint probability based on component probability is essential because it is difficult if not impossible to get experimental failure rate information on large equipments under widely different conditions whereas it is somewhat easier to get reliability information on components which may

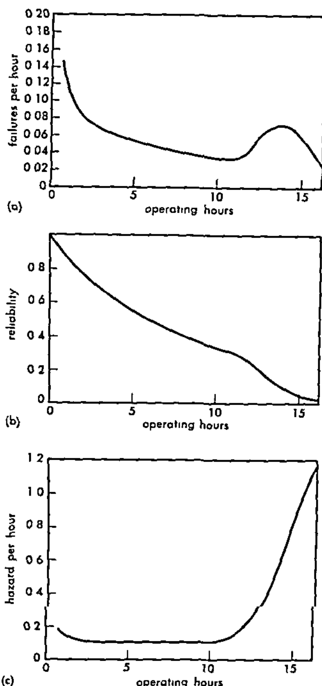


Fig. 1 Relationship between corresponding curves of (a) failure rate, (b) reliability, and (c) hazard rate. (From K. Henney, ed., Reliability Factors for Ground Electronic Equipment, McGraw-Hill, 1956)

be common to many different equipments and may be in fact duplicated many times over in a single equipment. This is especially true of electronic components such as resistors, tubes, and capacitors, which are used many times over in the same equipment and are used in many types of equipment.

Joint probability is the product of the individual probabilities. This definition assumes an independent relationship of the individual probabilities. Such independence is often not achieved in real situations for two reasons. First, because components of the same type often come from a common source, manufacturing or other considerations common to all the similar components may influence the reliability thereof. Secondly, in any equipment the functional interdependence of components cannot be overlooked. The failure of one device may influence the failure of an adjacent device due to load transferral, the *influencing of the immediate environment*, and many other factors.

Series reliability. A common case of joint probability is the series reliability case, where the failure of any individual component will cause the equipment to fail. Here the equipment reliability is the product of the reliability of the individual components. For the special case where all components have equal reliability, the reliability is then given by the reliability of the component raised to a power which corresponds to the number of components (see Fig. 2).

Parallel reliability. Redundancy can be introduced into a system so that in the event of the failure of a single component, a duplicate or paralleled component takes over the function of the failed part. Thus the probability that one of two paralleled components will survive is the sum of the probabilities of three possible favorable outcomes: failure of neither component A nor B, failure of A but not B, and failure of B but not A (see Fig. 3).

Reliability predictions. Reliability predictions are based entirely on experimental data. The results of the analysis of the experimental data predict a statistical conclusion which cannot be relied upon to apply to any particular element. Furthermore the complexity of the equipment greatly complicates the collection of data and analysis of probability prediction.

The scale of the experiments upon which the probability is based is important. A small number of experiments leads to coarse criteria; in fact there is some limit on the number of experiments which can produce useful results; thus the statistical sample must be adequate (*see* STATISTICS). The concept of satisfactory performance is also relevant, since the scale of the experiment is related to the ability to predict the probability of a range of satisfactory performances between all good and all bad. In order to predict reliably the probability

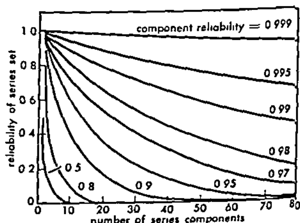


Fig. 2 Reliability of *n* series components of equal reliability. (From K. Henney, ed., Reliability Factors for Ground Electronic Equipment, McGraw-Hill, 1956)

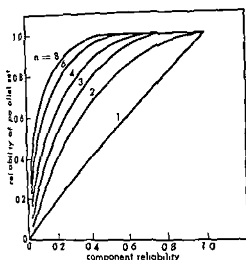


Fig 3 Rel ab lity of n parallel components of equal eliab l ty (From K Henney, ed , *Reliability Factors for Ground Electronic Equ pment*, McGraw Hill 1956)

of continuous variability as contrasted with dis crete off-on yes no black white decisions much data is needed

Engineering application Although the mathematical basis of reliability prediction is a science me reflection will demonstrate the difficulty of educing performance characteristics in a complex ystem to precise mathematical formulation In ad ition, complications arise because of the frequent nvolvement of a human operator whose operating

nces. Furthermore the evaluation of the reliability of even such apparently simple components as vac uum tubes is limited by inherent but inevitable lower order effects which theory cannot presently encompass Thus the experience and judgment of the engineer designing for or evaluating reliabil ity is vital

gress Older well known devices are more likely to be reliable than newly innovated techniques

Reliability cannot be adequately evaluated until after large scale production of a component But usually once large scale production is achieved some difficult to-alter decisions have been made which may in turn influence reliability Similarly in the case of redesign to effect improvements the risks of increased unreliability must be balanced against the improvements which may be achieved Only time and an experimental reliability program will indicate with finality the results

Reliability program. A sound reliability program for systems engineering includes the following

1 Complete awareness of the purpose of the equipment and the conditions of storage, transportation, and use in the field

2 Consideration of the systems aspects of the problem, the interplay between components, and their influence on reliability

3 Visualization of the operator's problem to make certain that the known limitations and capabilities of the human being are reflected in the design of the apparatus

4 Maintenance whenever possible, of conservative design and conservative use of components; knowledge of component performance under all conditions to avoid the misapplication of what would otherwise be perfectly good components

5 Emphasis on simplicity, because the reduction of the number of elements inherently increases the reliability

6 Inclusion of redundant and fail safe features in the design so that the failure of certain elements of the system does not affect its over all performance

7 Avoidance as much as possible of special parts and the preselection of parts use of inter changeable high volume production parts whose reliability can be more definitively evaluated

8 Recognition that maintenance often will be performed under difficult conditions and with limited facilities anticipation of the test equipment necessary to provide some guaranty of satisfactory performance

9 Selection of proper manufacturing methods with adequate quality control

10 Maintenance of a complete failure record with an explanation of the cause of the failures

11 Establishment of clear channels of information feedback from the operator in the field to the designer and manufacturer so that the cause of failure may be remedied the performance evaluated and maintenance techniques developed

See ENVIRONMENTAL TEST HUMAN ENGINEERING MAINTAINABILITY OF EQUIPMENT, MINIATURIZATION OF EQUIPMENT *see also* PILOT PRODUCTION, QUALITY CONTROL [R W M]

Bibliography K Henney (ed) *Reliability Factors for Ground Electronic Equipment*, 1956

Reluctance

A property of a magnetic circuit analogous to resistance in an electric circuit

Every line of magnetic flux is a closed path (*see* MAGNETIC CIRCUITS, MAGNETIC FLUX) Whenever the flux is largely confined to a well defined closed path there is a magnetic circuit That part of the flux that departs from the path is called flux leakage

For any closed path of length l in a magnetic field H the line integral of $H \cos \alpha \, dl$ around the path is the magnetomotive force (mmf) of the path

$$\text{mmf} = \oint H \cos \alpha \, dl$$

where α is the angle between H and the path (*see* MAGNETOMOTIVE FORCE) If the path encloses N

conductors, each with current I,

$$\text{mmf} = \oint H \cos \alpha \, dl = NI$$

Consider the closely wound toroid shown in the figure. For this arrangement of currents the magnetic field is almost entirely within the toroidal coil, and there the flux density or magnetic induction B is

$$B = \mu \frac{NI}{l}$$

where l is the mean circumference of the toroid and μ is the permeability. The flux Φ within the toroid of cross sectional area A is

$$\Phi = BA = \frac{\mu A}{l} NI$$

or

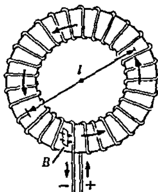
$$\Phi = \frac{NI}{l/\mu A} = \frac{\text{mmf}}{l/\mu A} = \frac{\text{mmf}}{\mathcal{R}}$$

This equation is similar in form to the equation for the electric circuit although nothing actually flows in the magnetic circuit. The factor $l/\mu A$ is called the reluctance $\mathcal{R}$ of the magnetic circuit. The reluctance is not constant because the permeability μ varies with changing flux density (*see* INDUCTION, MAGNETIC, PERMEABILITY, MAGNETIC). From the defining equation for reluctance it is seen that when the mmf is in ampere turns and the flux is in webers, the unit of reluctance is the ampere turn/weber.

Reluctances in series For the simple toroid, all parts of the magnetic circuit have the same μ and the same A. More complicated circuits may include parts that differ in permeability, in cross section, or in both. Suppose a small gap were cut in the core of the toroid. The flux would fringe out at the gap but as a rough approximation the area of the gap may be considered the same as that of the core. The magnetic path then has two parts, the core of length l_1 and reluctance $l_1/\mu_1 A$, and the air gap of length l_2 and reluctance $l_2/\mu_2 A$. Since the same flux is in both core and gap this is considered as a series circuit and

$$\mathcal{R} = \mathcal{R}_1 + \mathcal{R}_2 = \frac{l_1}{\mu_1 A} + \frac{l_2}{\mu_2 A}$$

Since the relative permeability of the ferromagnetic core is several hundred or even several thousand times that of air, the reluctance of the short gap may be much greater than that of the much longer core.



A toroidal coil

For any combination of paths in series, $\mathcal{R} = \Sigma l/\mu A$. Then

$$\Phi = \frac{\text{mmf}}{\Sigma \mathcal{R}} = \frac{\text{mmf}}{\Sigma l/\mu A}$$

Reluctances in parallel If the flux divides in part of the circuit there is a parallel magnetic circuit and the reluctance of the circuit has the same relation to the reluctances of the parts as has the analogous electric resistance. For the parallel circuit

$$\frac{1}{\mathcal{R}} = \frac{1}{\mathcal{R}_1} + \frac{1}{\mathcal{R}_2} + \cdots$$

See RELUCTANCE MOTOR. [K.V.M.]

Reluctance motor

A synchronous motor which starts as an induction motor and upon nearing full speed, locks into step with the rotating field and runs at synchronous speed. The stator and stator windings are similar to those of an induction motor. The rotor is of squirrel cage construction, to allow induction motor starting, and has salient pole projections which provide synchronous operation at full speed. The reluctance motor is built only in small sizes where low cost and simplicity are mandatory and where efficiency is of little concern. It can be polyphase, but is usually a single-phase motor with a split phase or capacitor winding for starting. *See* SYNCHRONOUS MOTOR; *see also* INDUCTION MOTOR. [L.V.B.]

Remote-control system

A system in which there is an appreciable distance separating the controlled quantity and the controlling quantity (*see* OPEN LOOP CONTROL SYSTEM). A familiar example of a remote control system is the telephone system. A voice at one end of

tity.

There are three essential components in a remote-control system: a controlling quantity, a transmission medium, and a controlled quantity. Three examples of remote-control systems are as follows.

1. Master-slave manipulator. This electromechanical device is used for handling radioisotopes from a safe distance. The controlling quantity is the hand motion of the operator, the transmission medium is a mechanical linkage system, and the controlled quantity is the position of the isotope container.

2. Telemetering system. A system for transmitting a measured quantity to a remote point is called a telemetering system. Telemetering systems may employ mechanical linkages, electric circuits, or radio waves as the transmission medium. *See* TELEMETERING.

3. Television. Television is used for transmitting a transient scene or picture from one location to

trolled quantity is the amount and position of light emitted from the face of a cathode ray tube *See* CONTROL SYSTEMS [J G TR]

Rennin

An enzymatic protein used for coagulating milk case n in cheese making It is also related to the cheese ripening process through its proteolytic ac tivity *See* CHEESE ENZYME.

Preparation Rennin is secured from rennet In preparing rennet extract, only milk fed calves are used. After butchering the fourth stomach *or abomasum is removed and freed of its food content* The stomachs are dry salted frozen and shipped in wooden barrels At the rennet factory the stom achs are washed freed of salt, and scraped to re move surface fat They are then stretched on racks which are wheeled into drying tunnels where a major portion of the moisture is removed The dried stomachs are placed in cold storage prior to being ground and mixed with ground excelsior in large vats A brine solution is continuously circulated through the skins until extraction of the rennin is complete.

Turb dity is an indication of decomposition The combined effects of low 4 4°C storage temperature and the presence of salt (approximately 15%)

is also made in the form of a powder or paste

Act on in milk The main protein in milk is case n and it exists as *calcium caseinate (see* CA SE N MILK) It is a complex mixture of at least

is that the enzyme promotes an unfolding of the polypeptide chains of the protein molecules expos ing reacti e groups These in turn are attracted to each other and become cross-linked forming a net work structure This action proceeds *in much the* same manner as in the formation of a polymer A clot, or coagulum is produced when the polymeriza tion has been carried far enough to produce a gel See FOOD ENGINEERING PROTEIN [P H T]

Repeater, synchro

A term appl ed *to* a class of electromechanical devices called self synchronous repeaters or syn chros t is also applied *to a class of remote in* d cat ng systems in which synchros are employed to transm t information *from one point to another* For a d scuss on of remote indicating systems *see* REMOTE CONTROL SYSTEM TELEMETERING

Synchros are electromechanical transducers wh ch convert a mechanical position into *an equiva* lent set of electrical voltages or transform from a set of voltages to an equivalent mechanical posi tion When the stator windings of two synchros are connected together and the rotors are energized by the same alternating current source a movement of one rotor will cause *the other rotor to move an* equal amount This operating feature makes syn chros useful as remote indicators of position that is *the driven synchro* repeats the position of the driving synchro *The rotor of the driving synchro* (or generator) is mechanically coupled to the de vice whose position is to be indicated remotely The driven synchro (or motor) rotor usually posi tions *the indicator needle of a dial For synchro* operation and types of synchros *see* SYNCHRO

A synchro repeater system which has been used *extensively* to give a remote indication of the posi *tion of naval guns* is shown schematically in Fig 1 In this system the generator rotor is mechanically fastened to the gun mount and the motor rotor is mechanically fastened to a dial indicator in the control room If there is any difference between the angles of the two rotors (θ and θ) currents flow between the two stator windings and cause the motor rotor to move in such a direction that ulti mately $\theta = \theta$ Using a system of this type the mechanical position of any shaft may be transmitted over a much greater distance than may be achieved by mechanical linkages

The synchro system shown in Fig 2 may be used to give a remote indication of the sum or differ ence of the angular positions of two mechanical shafts The addition of the differential generator

tial generator rotor before these voltages are ap *plied to the motor stator terminals* (S_1 S''_2 S_3) Depending upon the way in which the dif ferential generator rotor terminals (R'_1 R'_2 R'_3) are connected to the motor stator terminals (S_1 S_2 S_3) *the angle* θ_3 *of the motor rotor is the* sum or the difference of the angles θ_1 and θ_2

There are many other possible combinations of synchros which may be used to control a mechani *cal indicator from one remote location* or from a number of remote locations It is also possible to *control* a number of mechanical outputs from a

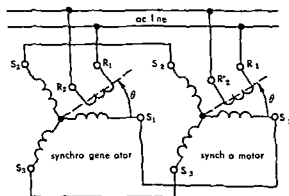


Fig 1 Remote nd cating system w th s ngle input

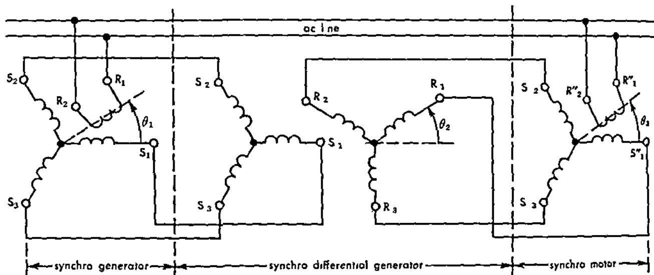


Fig 2 Remote indicating system with two inputs

single remote location In addition to their applications in remote indicating systems synchros are widely used as electromechanical transducers in feedback control systems *See* SERVOMECHANISM CONTROL SYSTEM [J G TR]

Bibliography W R Ahrendt *Servomechanism Practice* 1954 H Lauer R Lesnick and L E Matson *Servomechanism Fundamentals* 1947 J G Truxal (ed) *Control Engineers Handbook* 1958

Reproduction, animal

The formation of new individuals which may occur by asexual or sexual methods In the asexual methods which occur mainly among the lower animals the offspring are derived from a single individual Sexual methods are general throughout the animal kingdom with offspring ordinarily derived from the paired union of special cells the gametes of two individuals Basic to all processes of reproduction is the origin of the new individual from one or more living cells of the parent or parents There has been no acceptable demonstration of the origin of a new living organism from other material than preexisting living cells (*see* LIFE ORIGIN OF)

Asexual reproduction Asexual processes of reproduction include binary fission multiple fission fragmentation budding and polyembryony Among the protozoa and lower metazoa these are common methods of reproduction However the last mentioned process can occur in mammals including man

Binary fission involves an equal or nearly equal longitudinal or transverse splitting of the body of the parent into two parts each of which grows to parental size and form This method of reproduction occurs regularly among protozoans in which it is essentially the process of cell division with complete separation of the daughter cells To a limited extent binary fission may be observed among metazoans such as sea anemones as longitudinal fission and among planarians as transverse fission

Multiple fission schizogony or sporulation produces several new individuals from a single parent. It is common among the Sporozoa such as the malarial parasite which form cystlike structures containing many cells each of which gives rise to a new individual The cells of the cyst arise from a series of divisions of the nucleus which is followed later by cytoplasmic divisions of the original cell (*see* SPOROZOA)

Fragmentation is a form of fission occurring in some metazoans especially the Platyhelminthes or flatworms the Nemertinea or ribbon worms and the Annelida or segmented worms the parent worm breaks up into a number of parts each of which regenerates missing structures to form a whole organism It occurs also in certain starfish as *Linckia* in which single arms may pinch off and regenerate a complete animal

Budding is a form of asexual reproduction in which the new individual arises from a relatively small mass of cells that initially forms a growth or bud on the parental body The bud may assume parental form either before separation from the body of the parent as in external budding or afterward as in internal budding External budding is common among sponges coelenterates (Fig 1) bryozoans flatworms and tunicates Among certain of the coelenterates such as the colonial hydroid *Obelia* buds give rise to medusae or jellyfish rather than to the parental type polyp The medusae represent the sexual generation. They are free-swimming and of separate sexes producing eggs and sperm respectively Upon fertilization of the eggs the asexual polyp type individual develops (*see* METAGENESIS) Another example of asexual individuals budding sexual individuals is found in the cestodes or tapeworms Here the head or scolex by which the animal is attached to the host tissue produces a series of segments termed proglottids each of which is a sexual individual This phenomenon is also known as strobilization In some species of sponges coelenterates bryozoans

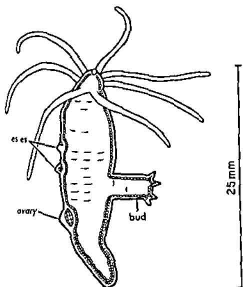


Fig 1 Budding in Hydra

and tunicates budding may occur without separation of the buds and thus lead to the formation of an organic colony accompanied in some cases by specialization of parts for particular functions such as in the Portuguese man-of war *Physalia*

Internal budding occurs among fresh water sponges and bryozoans In the sponges the internal buds, termed gemmules consist of groups of primitive cells surrounded by a dense capsule formed by the body wall If the parent animal dies as a result of desiccation or low temperature the cells of the gemmules can later be released and form new sponges In the bryozoans the similarly functioning buds are known as *statoblasts*

Polyembryony is a form of asexual reproduction occurring at an early developmental stage of a sexually produced embryo in which two or more off

. . .

...*Plus a* The embryo of this wasp subdivides so extensively that about 1500 individuals are formed.

In mammals, polyembryony is a regular feature of the development of armadillos The four offspring produced at a single birth in the nine banded armadillo *Dasypus novemcinctus* are identical quadruplets The quadrupling process occurs during the late blastocyst stage of development. In humans twins occur in about 1 1% of births, triplets in 0 012% and quadruplets in 0 00014% About one third of the twins are identical and have thus arisen by polyembryony probably occurring in a late blastocyst stage

Sexual reproduction Sexual reproduction in animals assumes various forms which may be classified under conjugation autogamy fertilization (syngamy) and parthenogenesis Basically the various processes all involve the occurrence of certain *special nuclear changes termed meiotic* divisions preliminary to the production of the new individual (see GAMETOGENESIS MEIOSIS)

Conjugation occurs principally among the ciliate protozoans such as *Paramecium* and involves a temporary union of two individuals during which each is ' fertilized by a micronucleus from the *other In this process the macronucleus of each* conjugant breaks down and the micronucleus undergoes two meiotic divisions to form four nuclei of which three degenerate The fourth nucleus divides again and each conjugant transfers one of these micronuclei to its partner where it fuses with the stationary micronucleus The conjugants *then separate and the fusion micronucleus in each* divides three times Of the eight nuclei four form macronuclei and three degenerate The exconjugants then divide twice, along with mitotic division of the micronuclei so that each of the four cells obtains one macro- and one micronucleus The parental state is thus restored and further reproduction occurs by simple fission The ability to conjugate again is not attained until after a large number *of divisions Conjugation does not ordinarily* occur within clones which are organisms derived by mitotic division from a single individual Strains that are capable of conjugating with one another are designated mating types rather than males or

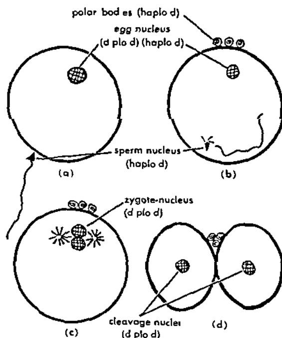


Fig 2 (a-d) Diagram of fertilization in animals In different species the sperm enters at one or another of the stages between a and b during which the two meiotic divisions of the egg occur

females, because there are no correlated morphological distinctions between them, and moreover, a particular species may have several mating types that are interfertile and thus represent more than two sexes

In autogamy, the nuclear changes described for conjugation take place, but since there is no mating there is no transfer of micronuclei Instead, the prospective migratory micronucleus reunites with the stationary one The process may be considered related to parthenogenesis

Fertilization, or syngamy, comprises a series of events in which two cells, the gametes, fuse and their nuclei, which had previously undergone meiotic divisions, fuse In metazoans, the gametes are of two morphologically distinct types, spermatozoa, or *microgametes*, and *eggs*, also called ova or macrogametes These types are produced by male and female animals respectively, but in some cases both may be produced by a single, hermaphroditic, individual The nucleus of the spermatozoon has half the number of chromosomes characteristic of the ordinary (somatic) cells of the animal The nucleus of the ripe egg in some animals, for instance, coelenterates and echinoderms, also has attained this haploid condition, but in most species of animals it is at an early stage of the meiotic divisions when ready for fertilization In the latter situation, the meiotic divisions of the egg, characterized by formation of small nonfunctional cells termed polar bodies, are completed after sperm entry, whereupon the haploid egg nucleus fuses with the haploid sperm nucleus Fertilization thus produces a zygote with the diploid chromosome number typical of the somatic cells of the species (23 pairs in humans) and this is maintained during the ensuing cell divisions (Fig 2)

In many protozoans, such as the flagellate *Polytoma*, the fusing gametes are not visibly different from one another nor from the ordinary organism They are termed isogametes In others they may differ (anisogametes) usually in size as in *Chlamydomonas braunii*, in some cases (*C coccifera*) one gamete is motile and the other not

Some metazoans are able to produce functional gametes while still in a larval condition Reproduction of this type, termed neoteny, is exhibited in amphibians by the axolotl, and also in various insects The Larvaceae among the tunicates are considered to represent persistent larvae reproducing neotenically

Hermaphroditism is rare among the vertebrates, but is found among most groups of invertebrates Despite the production of eggs and sperm by the same individual, most hermaphroditic animals reproduce by cross fertilization In some cases this is due to differences in time of ripening of the eggs and sperm In others with internal fertilization crossing is assured by differences in location of sperm and egg ducts In ascidians there is a physiological block to the union of the spermatozoon with an egg from the same individual The latter situation illustrates that the interaction of egg and sperm in fertilization is not only generally species-specific but may also be individual specific

Parthenogenesis is the development of the egg without fertilization by a spermatozoon It is listed as a form of sexual reproduction because it involves development from a gamete Rotifers, crustaceans, and insects are the principal groups in which it occurs naturally It has also been induced to occur (artificial parthenogenesis) in species from all the major phyla by various kinds of chemical or physical treatment of the unfertilized egg Even in mammals, several adult rabbits have reportedly been thus produced *See* EMBRYOLOGY, EXPERIMENTAL

The honey bee provides a classic example of natural parthenogenesis The males, or drones, all develop from unfertilized eggs and are haploid The females, or workers and queen, arise from fertilized eggs and are diploid Since the queen is inseminated only once, during the nuptial flight, she stores the sperm during her egg laying life, which is 5 years or more, and can evidently permit or prevent them from fertilizing the eggs that are laid

In certain animals periods of parthenogenesis may alternate with fertilization Thus aphids produce parthenogenetic females during part of the year, but as winter approaches males appear, whereupon fertilization ensues The fertilized eggs hatch out into females the following spring

Among certain of the Lepidoptera, like the gypsy moth *Lymantria*, parthenogenesis occurs at times when there appears to be a scarcity of males and the unfertilized eggs can form males or females There is also sporadic occurrence of parthenogenesis, recently reported, in chickens and turkeys Mostly the embryos are abnormal and die at an early stage, but in the turkeys many develop quite far, and a few have been reared to adults

Parthenogenetic development may also occur in eggs that are produced during a larval stage, as in the gallfly *Miaster* This condition is termed paedogenesis and is the parallel of neoteny *See* ESTRUS, OOGENESIS, OVUM, SPERM CELL, SPERMATOGENESIS [A TY]

Reproduction, plant

Plants produce new individuals either asexually or by sexual reproduction

Asexual reproduction. In simple one-celled plants, asexual reproduction is accomplished by division of the vegetative cell and the subsequent growth of the two nearly equal daughter cells which form new individuals (Fig 1) This is known as simple (binary) fission Another method known as budding or gemmation occurs in the unicellular yeast plant In this plant a small protuberance, which later becomes detached by constriction emerges from the cell The new cell thus formed may develop directly into a new plant, or it may undergo division and give rise to more new individuals In some liverworts and mosses, the plants produce small multicellular structures called gemmae, which, when detached, may grow, forming new

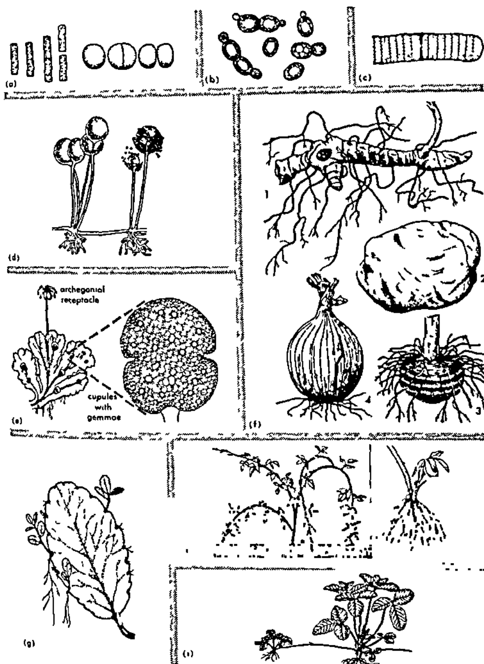


Fig 1 Types of asexual reproduction (a) Reproduction in bacteria fission in a bacillus and in a coccus (from H J Fuller and O Tippo, *College Botany*, Holt, 1949), (b) yeast cells in the living condition showing reproduction by budding, or gemmation (from A W Haupt, *Plant Morphology*, McGraw Hill, 1953), (c) species of *Oscillatoria* showing cellular structure of the filament highly magnified The lens-shaped object near the center is a gelatinous separation disk The filament breaks (fragments) readily at this point (from F W Emerson, *Basic Botany*, 2d ed, McGraw-Hill, 1954), (d) fungus developing sporangia, two with spores (from F W Emerson, *Basic Botany*, 2d ed, McGraw Hill, 1954), (e) liverwort showing cupules with gemmae (from H B Hill, L O Overholts, H W Popp, *Botany A Textbook for Colleges*, 2d ed, McGraw Hill, 1950), gemma of liverwort greatly enlarged (from G M Smith, *Cryptogamic Botany*, vol 2, 2d ed, McGraw Hill, 1955), (f) types of underground stems 1, rhizome of solomon's seal, 2, tuber of potato, 3, corm of jack in the-pulpit, 4, bulb of onion (from E L Core, *Plant Taxonomy*, Prentice-Hall, 1955), (g) leaf of *Bryophyllum* showing new shoots (plantlets) developed from adventitious buds in the notches in the leaf margin (from J B Hill, L O Overholts, H W Popp, *Botany A Textbook for Colleges*, 2d ed, McGraw Hill, 1950), (h) tip-layering of black raspberry (from E N Transeau, H C Sampson, L H Tiffany, *Textbook of Botany*, rev ed, Harper, 1953); (i) runner of strawberry with a young plant at the tip (E L. Core, *Plant Taxonomy*, Prentice-Hall, 1955)

plants Still another method occurring in many kinds of plants is effected by means of spores, which are usually formed by multiple division of an ordinary vegetative cell or in a specialized organ (sporangium) A sporangium may produce numerous spores and each one may produce a new plant Asexual reproduction also takes place by fragmentation in which a part of the vegetative plant becomes detached and develops into a new plant Many of the higher plants have underground stems, such as rhizomes, tubers, bulbs, or corms, which are efficient agents in vegetative or asexual propagation Offsets or sprouts developing at the base of stems plantlets arising asexually in the margins of leaves as in *Bryophyllum*, runners as those of strawberry arching stems which take root at the tips as in blackberries and raspberries, detached buds of certain aquatic plants, and other vegetative parts are of great importance in plant propagation In some instances these produce large populations All the plants derived from one original individual by repeated vegetative multiplication constitute a clone

In propagating plants for his own uses, man makes large use of vegetative reproduction Some plants such as horseradish and pineapple produce few or no seeds, and the seeds of others do not "come true", that is the new plants are, in some respects unlike the parents In all such cases successful propagation can be accomplished only by means of vegetative multiplication The grower often uses cuttings (parts of stems roots or even entire leaves) which are planted in soil where an adventitious root system develops and the cutting becomes an independent plant (*see* STEM CUTTINGS) Compared with propagation by seed this method is more dependable and produces plants in less time To propagate varieties of peaches, oranges and other fruits a bud of the desired variety is carefully cut from the twig and inserted in a slit in the stem of a developing seedling or in the branch of a mature plant (Fig 2) The bud may then grow into a tree or branch which will bear fruit precisely like that of the plant from which the bud was taken The trees of the Temple variety of orange have developed from Temple buds grown on inferior sour orange seedlings (*see* BUDDING) Grafting is essentially the same as budding except that twigs are used instead of buds, these are inserted in the cut end of a stem or branch The twig (scion) is carefully placed with its vascular cambium in contact with that of the stem (stock) so that a growth union may be formed Grafting and budding ensure the preservation of a given variety of fruit and by employing either one of these methods a single tree can be made to bear a dozen or more different varieties of fruit depending on the number of scions or buds used Grafting is also practiced to acclimate certain species of plants to strange environments, for example, grafting of plums onto the stocks of peach trees because plum trees grow poorly in sandy soils in which roots of peach flourish In Europe, grafting is done to check or eliminate certain parasites European grapes are seriously injured by a species of root louse which does not feed on the roots of American grapes. To grow these plants without damage by root lice the European grapes are grafted onto American grape stocks (*see* GRAFTING OF PLANTS)

Sexual reproduction Sexual reproduction appeared early in the evolution of the plant kingdom In this kind of reproduction, specialized sex cells called gametes fuse forming a zygote which has the potential to produce a new individual In some lower plants such as the alga *Spirogyra*, two protoplasts of ordinary vegetative cells may serve as gametes and unite but in others the protoplast undergoes multiple fission producing many motile cells If these are relatively large and few in number function as gametes fusing in pairs to form zygotes When these gametes are similar in size ap

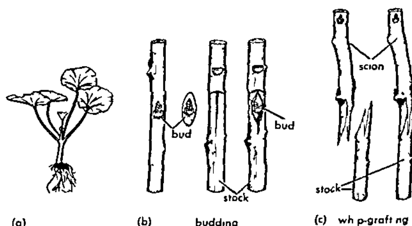


Fig 2 Types of asexual reproduction (a) Cutting from a geranium plant that has been placed in damp sand, showing development of adventitious roots from the cut end (from E W Sinnott and K S Wilson Botany Principles and Problems, 5th ed, McGraw Hill 1955) (b) budding and (c) grafting (from H J Fuller and O Tippo, College Botany, Holt, 1949)

pearance, and structure they are called isogametes, when morphologically different, they are heterogametes Therefore, fusion of isogametes is known as isogamy, and that of heterogametes as heterogamy In heterogamy, the larger, nonmotile gamete (female) is the egg and the smaller, motile gamete (male) is the sperm In thalloid plants, if the plant (thallus) produces both kinds of gametes, it is homothallic; if only one kind, it is heterothallic (*see* THALLOPHYTA).

Early in the course of evolution, certain cells of plants became specialized as sex organs to produce differentiated gametes The male organ producing numerous motile gametes (sperms) is called the

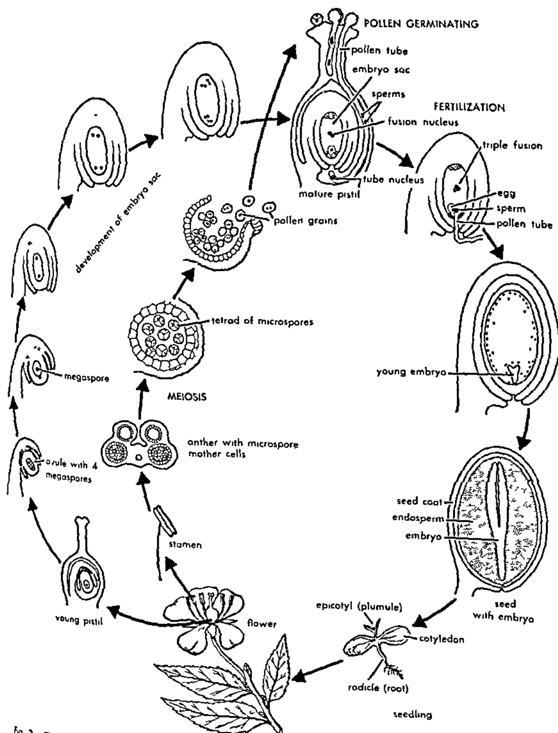


Fig 3 The life cycle of an angiosperm (E W Sinnott and K S Wilson, *Botany Principles and Problems*, 5th ed., McGraw-Hill, 1955)

antheridium, and the female organ usually producing but one egg is the oogonium In more highly organized plants, the one celled oogonium is replaced by a more complicated, multicellular archegonium containing one egg The plant bearing antheridia alone is the male gametophyte, and the one with archegonia is the female gametophyte Eventually, the seed plants appeared at the top of the evolutionary scale bearing flowers with stamens and pistils The stamens bear anthers (microsporangia) which produce the pollen grains (microspores) comparable to the asexual spores found in lower plants In the same flower, or in another, there is a pistil composed usually of an ovary, a style, and a stigma The ovary contains one or more ovules (megasporangia), in which the megaspores are formed One of these will develop into the embryo sac (female gametophyte) The pollen grain will germinate only on the stigma of a flower The transfer of pollen to the stigma (pollination) is accomplished by various agencies such as gravitation, wind, insects, birds, and other animals *See* EMBRYOPHYTA, FLOWER (BOTANY)

Pollination. Transfer of pollen from an anther to the stigma of the same flower or of another flower on the same plant is called self pollination When this transfer is from the anthers of one plant to the stigmas of another plant, it is called cross pollination In sexual reproduction, pollination is of significance only when it is followed by fertilization (union of sex cells) When many kinds of plants are blooming simultaneously promiscuous cross pollination occurs, but cross-fertilization occurs only between closely related plants, others are cross-sterile If fertilization follows self pollination, the plant is self fertile If fertilization occurs only after cross pollination, the plant is self sterile In some plants, such as grapes and pears, cross pollination usually occurs but if this fails, some fruit and seed of inferior quality often develop as a result of self pollination

The pollen grain (microspore) resting on a receptive stigma germinates giving rise to the pollen tube (male gametophyte) which elongates and extends through the style into the ovary where it reaches the ovule The nucleus of the microspore has already divided producing the tube nucleus and the generative nucleus The generative nucleus likewise divides forming the two male cells or sperms Within the ovule the germinating megaspore has developed into the embryo sac which varies considerably in different species of plants However, the embryo sac often contains eight nuclei of which two unite to form the fusion nucleus while another becomes the egg, these are the nuclei of greatest importance in reproduction

Fertilization. The pollen tube penetrates the ovule and enters the embryo sac where it swells and bursts releasing the two sperms The tube nucleus degenerates One of the sperms unites with the fusion nucleus forming the so-called triple fusion (endosperm) nucleus which divides and grows into the endosperm, a food accumulating tissue. The other sperm unites with the egg forming a zygote which may soon begin to grow at the expense of the endosperm, giving rise to the embryo (Fig 3)

Pollination and fertilization may affect structures outside the embryo and endosperm resulting in abscission of the pistil, or affecting chemical composition, color, and time of ripening of the fruit. Such influence is called ectogony In triple fusion genetic factors may be introduced by the male gamete which are responsible for certain characteristics of the endosperm such as color, form, shape, and chemical composition Appearance of these characteristics is known as xenia

Seed development. As the embryo and endosperm are developing, the ovule enlarges, its integuments become modified into protective coverings (seed coats), and the seed is fully formed As the embryo grows, it develops definite, specialized parts the hypocotyl, the first part to emerge when the seed germinates, gives rise to the primary root, the plumule, a little bud from which the aerial shoot develops, and the cotyledons (seed leaves) concerned with food utilization and food manufacture if exposed to light (*see* PHOTOSYNTHESIS) The endosperm may occupy space outside the embryo, or it may be digested, absorbed, and accumulated in the cells of the embryo In some plants the embryo develops immediately following fertilization but in others development is delayed for a time Under favorable conditions, some mature seeds will germinate immediately, whereas others will germinate only after a period of dormancy

Seed development is accompanied by changes in the ovary which enlarges and becomes the fruit. The fruits of plants display a wide range in form, size texture design and chemical composition, but in all cases fruits aid in dissemination of seeds *See* FRUIT (BOTANY), SEED (BOTANY)

Deviations. In the reproductive behavior of seed plants there are many deviations from the procedure as described Among these are parthenocarpy, or the formation of fruit without development of the ovules into seeds as in the banana polyembryony, or the production of two or more embryos within an ovule parthenogenesis or the development of an embryo from an unfertilized egg and the formation of perisperm, a food accumulating tissue resembling endosperm formed by enlargement and development of another structure the nucellus *See* PLANT [P D S]

Reproductive behavior

The different systems of behavior pattern by means of which in different types of animals, the sperm is brought to the egg, and by means of which the parental care of the resulting young is ensured The mere existence of eggs and sperm and of reproductive organs is not enough to guarantee that the eggs will reach the sperm There must also be, for each species a set of behavior patterns through which the male and female approach each other, and by their mutual reactions to each other, facilitate the movement of the sperm toward the egg

Sim larly there must exist patterns of parental be ha or *appropriate for each* type of animal by which the parent approaches the young and pro vides the necessary *care* The existence of milk producing glands in mother mammals or of heat producing arrangements *in the abdominal* skin of parent birds does not automatically ensure that the young mammal will get milk or *that the birds* eggs will be provided with the heat they require for incubation. The various patterns of reproduc tive behavior thus constitute one of the ways in which animals are *adapted* to their environment and have evolved under the influence of natural se lection just as have other aspects of structure func tion, and behavior The evolutionary influence of natural selection is to be seen in the fact that in each species, the repertoire of reproductive be havior forms a coherent and integrated system adapted to the characteristics of the environment the mate and the young Animals that are closely related to each other in terms of evolutionary origin tend to have similar behavior patterns whereas an mals more distantly related to each other will ordinar ly be more dissimilar in reproductive and other behavior Indeed it is often possible to de scribe and characterize a species genus or family of animals just as specifically and accurately by de scribing its characteristic behavior pattern as by describ ng its characteristic *structure* *See* SPECIA TION

Although the reproductive *behavior of inverte* brate an mals has interested many investigators and many aspects *of the reproductive behavior of* lower animals are relevant to the study of human behavior this discuss*ion is limited to the reproduc* tive behavior of nonhuman vertebrates and covers sexual behavior *and parental behavior*

SEXUAL BEHAVIOR

Sexual behavior means not only the behavior im

are usually restricted to definite locations In many animals, the place in which courtship and mating take place is an *area* defended by the male against the intrus on of other males of the same species Such a defended area the occurrence of which is associated with the reproductive phase of the ani mal's l fe is called a *territory* The classic exam ples of such breeding territories are to be found in connection with the *songbirds* When for example male snow buntings first arrive at their breeding lat tudes in spring (usually some days or weeks before the females) they at first stay together in flocks. After some days the males begin to sing and simultaneously to settle down in territories aban doning the flock which thus gradually disinte grates. Settled on its territory the male bunting selects one or more *prominent* perches and spends a great deal of his time singing from them The boundaries between territories are settled by fight ing in which actual physical encounters play a role but in which threatening postures and vocali zations may be much more prominent By the time the females arrive most of the territories are well established and the singing of the male birds serves to attract the females into them In the case of the snow bunting the territory is the area in which *almost all of the birds' activities* are carried out during the breeding season and in which they find *their food and later gather food for the young* It

sive source of food

and that this from the point of view of the evolu tion of territorial behavior is what determines the size of the territory There are other types of territories which do not have this function For ex ample many species of sea birds breed on shore in dense colonies from which the birds leave daily in large flocks to fish in neighboring waters In such colonies the defended territory may consist of a very small area around the nest perhaps no more than 3–4 ft in diameter Such a territory is defended just as vigorously and with just as highly devel oped ritualizations of behavior as the large terri tory of a songbird all of whose activities are con fined to it

Birds are not the only animals in which the de *fense of a territory plays a role in reproduction* Many male mammals fight during the breeding sea *son in defense of a specific area This may be a* fixed area as in the case of the wild relatives of *dogs and cats In other mammals such as deer* and many antelope which move in herds over large *areas the territory defended by a male may* be not a specific location but rather the space in the *neighborhood of his particular female(s)* Many mammals such as dogs hyenas bears and others mark the boundaries of their territories by urinat *ing on the ground or against rocks* or trees at the borders or by depositing there the secretions of special glands These scent flags may serve the same function in these animals which are highly sensitive to olfactory stimuli as do the singing and posturing of birds

Reproductive territory is also found in many spe cies of fish Some species defend specific areas whereas others defend the space around a particu lar object For example the female bitterling lays her eggs in a fresh water mussel and male bitter lings usually select a mussel and defend an area around it against other bitterlings

The characteristics of defended territories vary widely but all such reproductive territories seem to have in common the fact that they secure for the defending animal something essential for re production such as a mate food supply and nesting location

Courtship and pair formation Sexual relation ships among vertebrate animals range in different species from at one extreme temporary and pro-

miscuous copulations involving no individual recognition between particular males and females, to the other extreme of lifelong monogamous matings. These differences are not strictly characteristic of the major classifications of animals; rather a variety of types of mating system can be found in different species of the same class, and sometimes even of the same family, of animals. In some herd-living mammals such as the European red deer, a single male may have a harem consisting of several females, from which he drives away other males. Even more spectacular harems are to be found in some of the seals, among whom a single male may have 40–50 females. Other mammals such as wolves may have a more monogamous system of mating. In still other cases, as in many rodents, mating is casual and promiscuous. Similar variations are found among birds. In some species of birds, such as the European ruff and some species of grouse, the males gather together in an area called a lek, within which each male has a small area where other males do not trespass, and at the boundary of which he may fight or threaten neighboring males. This is, of course, a form of territory discussed previously. The females visit the lek and the males display to them as they pass through the area. Copulation takes place quite promiscuously, after which there is no further relationship between the male and the female; the female goes off to build her nest, lay her eggs, and rear her young alone.

At the other extreme, some species of geese may mate for many years, perhaps for life, forming very strong bonds between individual males and females, even migrating together and remaining together in the wintering area. A more common type of mating relationship among birds is an individual relationship lasting during a whole breeding season with, however, little or no evidence that the birds remain together in their winter habitat. Among fishes, too, different types of mating relationships may be found. Many species of cichlid fish form pairs which remain together throughout a breeding cycle and which appear to recognize each other individually. In other forms, such as guppies and swordtails, mating is promiscuous, and the animals do not stay together in pairs.

Egg fertilization. Fertilization of the eggs may occur in different groups of vertebrates either inside the body of the female or outside. In mammals fertilization is always internal, and copulation involves mutual reactions to each other of male and female which facilitate the insertion of the penis of the male into the vagina of the female and the release of sperm there. In birds, fertilization is also internal, but most species of birds lack a penis. Instead, both male and female have an opening called the cloaca on the underside of the body near the base of the tail. In copulation the male typically stands on the back of the female, balancing himself by fluttering his wings, and male and female tilt and adjust their tails so that the external openings are brought together for a brief contact, during which sperm is transferred from the male to the female. In some birds, such as many kinds of ducks, there is an erectile male organ slightly resembling the mammalian penis, which is everted from the cloaca during copulation. Internal fertilization is the mode in reptiles, in most groups of which the male possesses an erectile organ of copulation, the hemipenis, which is similar in function to the mammalian penis. In amphibians, fertilization of the eggs takes place without insertion of a copulatory organ by the male into the female. In frogs and toads, the male mounts the back of the female and clasps his forelegs tightly around her body. While engaged in this clasp, or amplexus, the female periodically releases groups of eggs, and the male simultaneously releases sperm which fertilize the eggs in the water. Mating patterns in most of the tailed amphibians, or salamanders, involve a curious form of internal fertilization. Following a more or less elaborate precopulatory courtship display, the male deposits a spermatophore, a small gelatinous stalk surmounted by a cap of sperm. The female, walking behind the male during this ceremony, enfolds the cap of the spermatophore between the lips of her cloaca, thus taking it into her body where fertilization is accomplished. This type of fertilization may take place either on land or in the water, in different species.

Among fishes, fertilization may be either internal or external. In live-bearing fishes, the anal fin is modified into a gonopodium, which is inserted into the genital opening of the female and serves as a guide for the sperm. In other groups of fish, such as the various species of salmon and trout, fertilization is entirely external; the male and female, swimming or lying side by side, simultaneously eject eggs and sperm, the fertilization taking place in the water. Although in this type of mating there is no physical contact between the male and the female, in many species there can be seen vigorous, highly synchronized movements, involving apparently synchronized increases in and relief from tension, highly reminiscent of sexual orgasm in mammals. See COPULATORY ORGAN.

Physiological regulation. Stimuli involved in courtship and sexual behavior may involve any sensory modality, and different species of animal differ widely with respect to the varieties of stimuli used in courtship. Observations of the sexual behavior of mammals in zoos and in nature clearly indicate that visual, auditory, olfactory, and tactual stimuli may all be effective in arousing sexual excitement and in the recognition of the mate. On the other hand, experiments on the common laboratory animals (rat, guinea pig, and cat), confined under admittedly rather unnatural conditions, show that effective copulation may occur in animals deprived of any of the sensory modalities, although not of all. Observations and experiments on birds show that both auditory and visual stimulation are effective in courtship. In some species of bird the male reacts differently and appropriately to stuffed and mounted specimens of males or of females of

his own species Female birds of many species are attracted by the courtship songs of the males

Visual auditory, tactual and chemical stimuli have all been shown to be effective in various of the lower vertebrates (fish amphibians reptiles), several of them sometimes being effective at different stages in the sexual behavior of the same species. For example the males of certain species of frog approach and clasp other frogs on the basis of their visual characteristics and then either

sound and has the body shape characteristic of an other male or of an egg laden female Female salamanders are sometimes stimulated to follow the male during the prefertilization courtship behavior by stimulation coming from the secretion of special glands in the male Many aspects of the courtship of fishes are influenced by visual stimulation provided by the mate Experiments with artificial models, for example show that a male three spined stickleback on its territory reacts differently to an approaching fish depending upon whether its color and shape are characteristic of a male or of a female

The effects of stimulation during courtship may either be direct immediate effects upon the behavioral reactions of the other animal or they may induce responses quite remote from the immediate behavior For example the characteristic posture, movements, and odor of a female rat in heat quickly induce male rats to follow her and to react to her in ways different from those in which they react to females not providing these stimuli The attack of a male flicker (woodpecker) against an other male intruding on his territory seems to be

stimuli provided by the other lies in the fact that behavioral interactions between members of a pair may have the effect of synchronizing their moods so that they are at the same level of intensity when they reach the part of the behavior pattern (for example, copulation) at which detailed synchronization of the behavior is of the greatest importance Many of the prespawning behavior patterns of fishes appear to have this synchronizing effect

A still more indirect but nevertheless very important, effect depends upon the fact that stimuli provided by the activities of one mate may actually stimulate changes in endocrine secretion in the other and thus alter its physiological condition in such a way as to bring on the next stage in the breeding cycle. For example the onset of puberty and the establishment of regular estrus cycles in female mice may be influenced by olfactory stimuli coming from the males. A female dove may be induced to lay an egg solely by seeing a male dove court her, even through a glass plate In general, the courtship behavior of male birds seems to influence the endocrine glands of females to secrete the hormones which in turn induce them to become sexually receptive Visual stimuli provided by other members of her species induce the female African mouth breeder (a cichlid fish) to dig nests and lay eggs Much evidence now (1959) available indicates that these psychosomatic influences of external stimuli on endocrine secretion may be of considerable importance in the regulation of the sexual cycles of many different kinds of animal In seasonal breeding animals the testes of the male and the ovaries of the female are small and physiologically quiescent during the nonbreeding part of the year, enlarging and becoming active during the breeding season In birds and other spring breeding animals the seasonal development of the sex glands is stimulated by the increasing length of days in the spring This was first shown in a dramatic experiment in which a group of male birds was exposed in the fall to increasing daylight by the use of additional artificial illumination, while another group was kept under normal conditions of decreasing daylight At the end of the treatment period during the winter the control group had as expected small inactive testes, whereas the light treated group had the large active testes normally characteristic of birds at the height of the breeding season The length of the day thus by its physiological effect upon the endocrine system determines the breeding season in temperate zone birds However the actual time of breeding within the season is subject to variation influenced by various other stimuli such as temperature changes the availability of appropriate habitats and stimuli from the mate Growth and activity of the sex glands like those of most of the other endocrine glands are under the influence of hormones coming from the pituitary gland The fact that the activities of the pituitary gland are in turn controlled in considerable detail by the brain provides the physiological background for the ability of all these external stimuli to influence the activity of the testes and ovaries

Direct evidence is available that hormones secreted by the testes and ovaries (gonads) play an important role in the development and occurrence of sexual behavior Removal of the gonads is generally followed by disappearance or sharp reduction of sexual behavior although the administration of sex hormones to animals whose gonads have been removed or to immature animals whose gonads have not yet begun to function is usually followed by the appearance of sexual behavior The abruptness with which sexual behavior disappears after removal of the gonads and the extent to which sexual behavior persists after such removal vary in different kinds of animals and under different conditions For example in some male mammals such as the domestic cat castration before puberty will prevent the development of sexual be-

havior If castration occurs after puberty, the ability to perform sexual behavior may die out quite quickly in animals that have had no sex experience before castration, whereas in animals that have had sexual experience, sexual behavior may persist for a considerable time In rats, which may be regarded as lower in the evolutionary scale than cats, individual experience does not seem to play the same role in determining the influence of sex hormones On the other hand, in the monkeys and apes, learned capabilities and drives influenced by learning appear to play an even greater role, and the immediate presence of sex hormones seems to be less critical than in the case of the cat There may thus be a general evolutionary trend toward increased significance of learning and decreased dependence upon the immediate influence of sex hormones in the organization of sexual behavior However, substantially different patterns may be found in different species at the same general level of evolution For example, social experience in early life seems to play a negligible role in the development of sexual behavior in laboratory rats, although it has a substantial effect in the guinea pig Further, *females in general tend to be more dependent upon* the presence of sex hormones for the performance of sexual behavior than do males of the same species Questions of this type have not been investigated in vertebrates other than the mammals

PARENTAL BEHAVIOR

Once the egg has been fertilized, the eggs or the young or both cannot survive unless the behavior of the parent is appropriate to meet the needs of the offspring during the period when they are dependent upon the parents The patterns of parental behavior in animals which bear live young are different from those which produce eggs requiring care before they hatch

Nest building. Nest building patterns in birds mammals, and fish are discussed in the following sections

Birds During the breeding season most species of birds build a nest and lay their eggs in it Birds' nests vary widely in structure and location Some species of weaver birds build massive, elaborately woven structures of grass, supported in the branches of trees A large number of pairs may cooperate in building such a structure, each pair having a hollowed out chamber of its own in the communal nest structure, with an individual opening leading to each chamber Most birds, however, build individual nests The American orioles build elaborately woven completely enclosed nests of grass, hanging from the branch of a tree Many other birds build open nests supported in the crotch of the branch of a tree, or among the twigs of a bush These may be deep-woven cups, sometimes lined with mud, like that of the American robin, or simple platforms of twigs, like the nests of most pigeons and doves Some birds characteristically nest in hollow trees, others, like woodpeckers, dig out nest chambers in the trunks of trees by their own efforts Many sea birds such as the terns build very simple nests, little more than a depression hollowed out in the sand, lined with a few sea shells. Some species of sea birds which breed on narrow ledges on the faces of ocean cliffs may build no nests at all laying their eggs directly on the rock

Either sex or both may participate in the building of the nest. There are a few species such as the phalaropes in which the male does all of the nest building However, by far the more common patterns are for the female to build the nest alone, as do many songbirds (thrushes, sparrows, blackbirds), or for both sexes to participate, as do doves, gulls, and others

Although nest building behavior is here included in parental behavior because the function of the nest is to provide a place in which the eggs can be incubated, the act of building a nest is frequently closely associated with, or even a part of sexual behavior The nest is usually built during the last few days before the first egg is laid, a period during which precopulatory behavior and copulation are at a maximum In many species of birds in which the female does all of the nest building, *the first nest building activity of the female* occurs simultaneously with the beginning of her sexual receptivity In some species such as the night heron both birds participate in the building of the nest, and the early stages of nest building are actually a part of the courtship activity In still other species such as many shore birds courtship behavior includes the scraping out of a nest hollow in the sand by movements of the feet and breast of the crouching male who may scrape out several such nests, in one of which the female lays her eggs

Mammals Although the nest building behavior of birds is closely related to the time of breeding reproductive nest building behavior in mammals ap-

mammals such as weasels and many rodents live in burrows which they dig underground These burrows may be lined with material collected by the animals such as leaves and twigs The building activity does not appear to be particularly related to the breeding season Similarly, field mice, squirrels and a number of other groups build globular nests of grass or twigs above the ground independently of the breeding cycle Chimpanzees build nests in trees in which they sleep, a new nest being built each night at a different location

Although little information is available about seasonal variations in nest building behavior in these wild mammals, laboratory mice and rats show well defined changes in nest building behavior associated with pregnancy and parturition They *engage in nest building* behavior to some extent when they are not breeding, but the amount of such activity increases sharply toward the end of pregnancy, with the result that a relatively massive well formed nest is usually built by about the time when the young appear Domestic rabbits build

nests of grass, which they line with hair plucked from their own bodies In these animals, the nest may be built just before parturition or just after,

there may be found variations in nest building behavior related to pregnancy and parturition similar to those now known for domestic types

Fishes Some fishes build rather elaborate nests, whereas others build none at all Salmon, for example simply lay their eggs on an appropriate gravelly bottom so that the eggs slip between the stones Jewel fish dig a shallow depression in the sand at the bottom of the stream or aquarium in which the eggs are deposited The males of some species such as the Siamese fighting fish catch in their mouths the eggs which the female has just released (and which the male has just fertilized), then blow a bubble of air which floats to the surface

secretion produced by a modified kidney

Incubation of eggs In lower vertebrates parental care of the eggs takes a variety of forms Many of these animals may have nothing to do with the eggs after they are laid This is true of many fishes and amphibians and some reptiles (for example tortoises) Other species belonging to the same classes may however care for the eggs in a variety of ways The male three spined stickleback remains near the nest during the period before the eggs hatch performing characteristic fanning movements with his fins These movements cause

the absence of the fanning movements In other species, such as some of the cichlid fishes this type

snakes remain in contact with their eggs during their development providing some heat (but not very much in view of the low heat production of these cold blooded animals)

In birds, which are warm blooded animals the eggs invariably require for the optimal development of the embryos a temperature higher than the environmental temperature The additional heat is provided, in almost all species of birds by the body of one or both parents The incubating parent sits on the eggs in doing so it erects the feathers of the lower abdominal region to expose an area of

the skin layers, and in some cases loss of the downy feathers which are present at other times of the year Such an area of modified skin which is adapted for the local transmission of body heat, is called an *incubation patch*. In doves, ducks, and some others, no incubation patches have been found, although incubation occurs in the same manner as in species with the patches

In some species of birds, including most song birds the female does all the incubating and the incubation patch is present only in the female In some families or groups of species, both males and females take part in incubation, and have incubation patches In these cases, the males and females may change places on the eggs at irregular short intervals of from 10 min to 2 hours (as do most song birds) or they may each sit at a particular time of the day (as do doves), or each may sit for several days at a time (as do some petrels) In a few species, such as the phalaropes only the male sits on the eggs

Aberrant patterns of incubation are to be found in several bird families Among the megapodes, or mound builders a family of birds found in Australia and the southwest Pacific, the males build large mounds of leaves grass, and other materials in which the females lay their eggs after which the eggs are covered with more plant material The incubation of the eggs is then provided for, not by heat from the birds' bodies but by the heat produced by the decaying substance of the mound In some species of this family the young, when they hatch receive absolutely no parental care The American cowbird and most species of Old World cuckoos are parasitic that is they build no nests of their own but lay their eggs (one to a nest) in nests of birds of other species leaving the eggs to be hatched and the young to be reared by the foster parents

Pregnancy and birth Pregnant mammals of many species tend to be quieter and more solitary than nonpregnant females The female chimpanzee is gentle and unaggressive in contrast to her usual behavior Pregnant cats are relatively inactive and spend an increased amount of time licking and grooming their bodies particularly the genital region During late pregnancy herd living mammals of some species such as the American elk tend to leave the herd finding a secluded place in which to give birth

Mammals of different families give birth in characteristically different positions Some, such as the giraffe deliver the young from a standing position Others such as the cat may assume a lying or a sitting position at different stages during the birth process Most rodents sit semiupright on the hind legs while giving birth The rabbit stands in a peculiarly crouched position so that the young are delivered forward to lie under the mother's body

Most mammals eagerly lick their genital areas, the young, and the fetal membranes during and after parturition The fetal membranes are usually torn free by the mother, who also bites through the umbilical cord In most mammals, the mother also eats the placenta, or afterbirth The eating of the afterbirth may be seen even in such species as the guinea pig, rabbit, and bison which are strictly vegetarian at all other times The licking of the young appears to be an important aspect of the establishment of mother young relationships, because in many species it helps to orient the mother and young to each other so that suckling can be established For example, if a domestic lamb is removed from its mother immediately after birth, washed, and returned some hours later, the mother and lamb may never establish a suckling relationship In addition, the maternal licking of the anogenital region of the young stimulates the first urinations and defecations by the young In many species this cannot occur without such stimulation In some mammals such as the camels and their relatives, and in many mammals which give birth in the water such as porpoises licking of the young and eating of the afterbirth do not occur

Care of the young. The parental behavior patterns in the care of the young bird and the mammal are considered in this section

Birds Young birds at the time of hatching from the egg may be one of two types, altricial young, hatched naked blind incapable of locomotion, and unable to get food for themselves, and precocial young, hatched covered with down, able to walk, the eyes fully developed, and often able to secure food for themselves Altricial young are brooded and fed by the parents for some time after hatching At first the young beg for food by movements which are not oriented to the parents, simply holding the head pointed straight up with the mouth open so that the parent can drop food into it They may often need to be fed 10 times or more in a single hour Most songbirds such as thrushes sparrows and flycatchers are of this type Precocial young on the other hand require brooding for only a very short time, and peck directly at their food from the first In some species with precocial young such as chickens and ducks the parents lead the young to the food, the young pecking for food at the ground just as do the parents In other cases such as gulls and shore birds the parents may bring the food to the young and either hold it in front of them or drop it before them so that the chicks can peck at it In species in which only one parent sits on the eggs, the same parent always broods the young However a parent which neither sits on the eggs nor broods the young may play a major role in feeding the young For example a common pattern among songbirds is that in which the female does all the incubation of the eggs and all the brooding of the young, whereas the male does most of the feeding In some other birds such as gulls, both male and female share in incubation, brooding, and feeding of the young Food may be carried to the young whole, in the claws of the parent, as by some hawks, it may be carried within the mouth of the parent, as by most songbirds, it may be swallowed by the parent, partly digested, and regurgitated for the young, as by gulls, or it may actually be produced as a secretion in the crop of the parent and regurgitated to the young, as by pigeons and doves and probably hummingbirds.

Mammals Mammals provide food for their newborn young by producing milk in the mammary glands, at the nipples of which the young suck At first, the initiative in establishing nursing appears to come from the mother at each episode, but as the young develop, the young take the initiative more and more often in approaching the mother Nursing occurs in various positions Cats and dogs nurse their young while lying on their sides, most of the herd living mammals nurse in a standing position, many rodents crouch over the young, monkeys and apes hold the young against the breast in their arms The frequency of nursing varies considerably Rabbits nurse their young only once a day, whereas young porpoises may suckle for a few seconds at a time every 15–30 min, day and night Mother seals of some species may spend a day or more ashore with their young and then several days at sea, during which time the young are not nursed The frequency of nursing gradually declines, the weaning of the young (that is, the transition from suckling to procuring solid food) taking place at different ages in different species Mice may be independent of the mother about 2 weeks after birth whereas some herd living animals (elks, bison) may still be suckling from the mother when they are 9 months or 1 year old, although at this age they will be getting most of their food by grazing

Physiological regulation. As in the case of sexual behavior the stimuli involved in the evocation of the patterns of parental behavior may be in any sensory modality and often differ in different species of animal For example the retrieving of displaced eggs into the nest in some species of lizards is apparently dependent upon the chemical characteristics of the eggs The regulation of the amount of fanning done by the male three-spined stickleback when guarding the eggs is partly influenced by the amount of carbon dioxide in the water, a higher concentration of carbon dioxide induces a greater amount of fanning activity Responses to the young in many species of fish which care for their offspring such as the mouth breeders are based on visual stimuli Responses of birds to their eggs are based on visual tactual and temperature stimuli If an egg is colored differently from the rest of the eggs in the nest birds of many species will refuse to sit on the nest in other species the birds will sit only after ejecting the odd egg, in still others a peculiarly colored egg may be incubated but not rolled back into the nest if it should fall out Finally, some species will not be at all disturbed by the strange egg If the egg of an incubating gull is maintained at an abnormally high or

low temperature by circulation of heated or cooled water through it the bird will *be noticeably* restless and disturbed indicating that temperature stimuli play a role in regulating the incubating behavior Reactions of birds toward their young may be based on visual and on auditory stimulation provided by the young Olfactory stimuli have not been shown to be important in regulating the behavior of birds.

Mammal mothers recognize their young and are stimulated to respond to them by various combinations of olfactory auditory visual and other stimuli. Some mice retrieve their young in response to supersonic cries Rats may retrieve young when stimulated by their smell sound or visually perceived movements. In some species of herd living mammals such as seals mothers appear to recognize their own young by their voices

*As noted previously in connection with the regu*lation of sexual behavior stimuli provided by the young may either cause immediate behavioral responses of the parent, or may indirectly cause changes in the parent s behavior by stimulating changes in their physiological condition The sight of eggs in the nest stimulates the parent bird to sit *on them* and the movements and sounds made by young birds stimulate the parents to feed them The mother rat is stimulated to retrieve and to nurse her young by their sight sound and smell In addition however the presence of eggs or of young stimulates more pervasive changes in the parents.

When doves sit on their eggs. their crops in

is known that the growth of the crop is induced by the hormone prolactin secreted by the pituitary gland this experiment demonstrates that stimuli provided by the eggs induce the secretion of prolactin by the parents pituitary glands Further when the newly hatched young of wild wrens are removed the parents begin the courtship and nest building which will *lead to* the production of a second brood about 1 week earlier than they would otherwise do This indicates that stimuli provided by the young inhibit the secretion of those pituitary hormones which induce the beginning of the new breeding cycle. *See* HORMONE.

Suckling stimuli provided by the young mammal induce the secretion *of* oxytocin a hormone of the posterior lobe of the pituitary gland The oxytocin in turn, when it reaches the mammary gland causes an increase in pressure in the milk secreting tubules, so that milk is actively squeezed out toward the nipples. Milk is thus actively expelled by the mammary gland and the mechanism of this ejection is through the secretion *of* a pituitary hormone in response to stimuli provided by the young If the young are removed from a mother mouse or rat her mammary glands will soon cease to secrete milk and she will stop behaving maternally toward young offered to her On the other hand it is possible to make lactating females continue lactation far beyond the normal weaning period by repeatedly replacing the growing litters with newborn young These and other experiments indicate that the stimulation of the nipples by the suckling *young induces the secretion of the* pituitary hormones which maintain milk production in the mammary gland The stimulus for the hormone secretion is actually the mechanical stimulation of the nipples the same effect can be produced by painting the nipples with turpentine instead of allowing young to suckle them *See* PITUITARY GLAND

Because parental behavior occurs at specific times during the reproductive cycle changes in endocrine secretion characteristic of the reproductive cycle should play a role in regulating such behavior Doves injected with progesterone a hormone normally secreted by the ovary may be induced to sit on eggs Laying hens are made broody (that is induced to sit on eggs) by injections of the pituitary hormone prolactin A combination of ovarian and pituitary hormones is responsible *for the development of the incubation* patch Prolactin which stimulates the growth of the crop and the secretion of crop milk in doves also induces these birds to regurgitate food to squabs (young doves) It thus appears that various aspects of the parental behavior of birds are in fact stimulated and regulated by the effects of endocrine secretions As in the case of hormone-induced sexual behavior the effects of hormones sometimes depend upon the previous experience of the animal Doves which have never had *breeding* experience may fail to feed squabs when injected with prolactin and inexperienced doves when injected with progesterone do not sit on eggs so often nor so quickly as do experienced birds

Nest building behavior is induced in mice by the injection of progesterone the ovarian hormone *which normally appears in the mouse s blood* in midpregnancy when intense nest building usually begins Nest building may also be induced by introducing newborn young mice in the cage In a normal cycle nest building is started during pregnancy because of the effects of progesterone and continued after parturition (when progesterone se*cretion stops*) by *the effects of the young* In rats nest building appears to be partly a temperature regulating device because it is greatly increased by keeping the animals in a cold environment and because removal of the thyroid glands (which results in interference with the temperature regulating mechanisms of the rat s body) results in an enormous increase *in nest building activity* which compensates for the now deficient temperature regulation by providing additional insulation for the animal s body When injected into rats prolactin the hormone which is partly responsible for milk production after parturition also increases the frequency of retrieving of the young

Hormonal bases of parental behavior have not been investigated in lower vertebrates *See* BEHAVIOR AND HEREDITY, PERIODICITY IN ORGANISMS, PROGESTERONE, PSYCHOLOGY, PHYSIOLOGICAL AND EXPERIMENTAL [D S L]

Bibliography F A Beach, *Hormones and Behavior*, 1948, P P Grasse (ed), *L'Instinct dans le comportement des animaux et de l'homme*, 1956, N R F Maier and T C Schneirla, *Principles of Animal Psychology*, 1935, W H Thorpe, *Learning and Instinct in Animals*, 1956, N Tinbergen, *Social Behavior in Animals*, 1953

Reproductive system

The structures concerned with the production of sex cells (gametes) and perpetuation of the species The embryology, comparative anatomy, histology, physiology, endocrinology, and biochemistry of this system are treated in this article

The reproductive function constitutes the only vertebrate physiological function that necessitates the existence of two morphologically different kinds of individuals in each animal species, the males and the females (sexual dimorphism)

The purpose of the reproductive function is fertilization, that is, the fusion of a male and a female sex cell produced by two distinct individuals In each sex the reproductive system comprises a sex gland or gonad which produces sex cells or gametes, and ducts which permit the passage of the gametes In some animals such as mammals, copulatory organs permit the male germ cells to be introduced into the female ducts and fertilization is internal, but in a great number of vertebrates, such as anuran amphibians or fishes, no copulatory organ exists and fertilization is external *See* OVARY, SEXUAL DIMORPHISM, TESTIS

EMBRYOLOGY

The embryology of the reproductive system is very similar among all vertebrates, with the exception of teleost fishes in which it is less specialized It proceeds by successive steps

Sex determination. At the time of fertilization the sex of each individual is genetically determined by the sex chromosomes contained in the gametes If both the male and the female germ cells have an identical X chromosome, the egg, upon fertilization, receives two X chromosomes (homozygous egg) If one parent germ cell contains an X and the other a Y chromosome, the egg cell receives an XY assortment (heterozygous egg) In mammals and in frogs, for instance, the male is the heterozygous XY sex In birds and several urodeles the condition is reversed and the female sex is the heterozygous XY sex

The fundamental importance of the genetic basis of sex determination must be emphasized before the embryology of the reproductive system is studied, because it should be considered as the first step in sexual differentiation of each individual *See* SEX DETERMINATION

Later during development, the gonads, the sex ducts, and corporal sex characters appear suc sively

Role of urinary system. Before more spe details are considered, it should be noted that development of the genital tract is intimately related with development of the urinary sys (Figs 1 and 2), and that the urinary system it is derived from successive kidney organs. first pair of kidneys, the pronephros of the e embryo, develops mainly in the future neck reg Its duct, the primitive ureter, reaches the clo Shortly afterwards the second pair of excret organs, the mesonephros, differentiates appr mately in the middle of the trunk, in connec with the primitive ureter The pronephros then rogresses and the primitive ureter becomes mesonephric duct, also called the Wolffian d In higher vertebrates (birds and mammals) mesonephros is later replaced by a third pair kidneys, the metanephros

Among the vertebrates, except for the tele fishes, the sex glands differentiate in connect with the mesonephros They retain these con tions, especially in the males in which the m nephros remains the adult kidney (Fig 8) adult male frog displays conditions which are s ilar to those found in the early bird or m malian embryo In both, the testicular tubules connected with mesonephric urinary tubules URINARY SYSTEM, UROGENITAL SYSTEM

Development of the gonads. The developm of the gonads is a progressive process, which m be divided into three main phases (1) the appe ance of a genital ridge, (2) organization of an differentiated gonadal anlage, and (3) sexual d ferentiation of this primordium

Genital ridge The genital ridge appears on t mesolateral side of the mesonephros as a thick

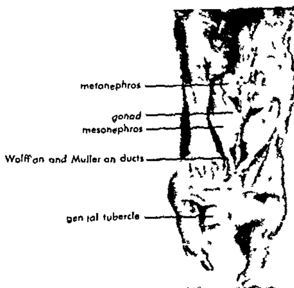


Fig 1 Dissection of a 19-day-old rabbit fetus sho ing the undifferentiated condition of the genital tra of a mammalian fetus

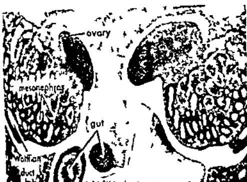


Fig 2. Section through the body of a 17-day-old rabbit fetus showing the ovaries at an early stage of differentiation located on the internal side of the mesonephros

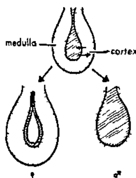


Fig 3 Schematic presentation of sex differentiation of amphibian gonads The indifferent anlage with its cortex and medulla may differentiate either in the female or in the male direction

or 2 3 cells thick This contrasts with the flat cells of the other parts of the mesonephric coelomic wall. Primordial germ cells the cells which will give rise to the germ cells, come in close contact with the germinal epithelium and even penetrate between its cells These primordial germ cells have an extragonadal origin

Extragonadal origin In the chick embryo the primordial germ cells are first *seen as large undif*ferentiated cells located in the germinal crescent a part of the extraembryonic area of the blastoderm situated anteriorly *to the embryo's future* head. They reach the level of the gonadal anlage by way of the blood stream (C Swift, 1914, D Simon 1958) In sauropsidians and *mammals* their extragonadal origin has also been established although their mode of migration toward the germinal epithelium remains uncertain (*ameboid* migration is often assumed)

In the frog the primitive germ cells have been traced from the very beginning *of the cleavage* of the egg by L. Bounoure At the time when the gonadal anlage differentiate they are disposed in more or less metameric clumps near the inner border of the mesonephros

Amphibians As soon as the germinal ridge becomes apparent, the germ cells protrude into the abdominal *cavity with the coelomic* epithelium The anlage is then invaded by clusters of cells coming from the mesonephric blastema which first exhibit a metameric disposition The gonadal anlage has then attained a stage known as the indifferent stage, which is the same irrespective of the genetic sex of the animal

The indifferent sex gland consists of the coelomic epithelium at the periphery, whose cells are flat and scattered and cover the voluminous primitive germ cells Some connective tissue surrounds the central mass of small cells which came from the mesonephric blastema The outer zone is known as the cortex, the inner part as the medulla (*E. Witschi, 1914*)

During the weeks preceding metamorphosis, the undifferentiated anlage may differentiate either as an ovary *or as a testis*

Ovarian organogenesis This process is characterized mainly by the development of the cortex The germ cells become surrounded by small follicle cells and begin to increase in size and to show some premeiotic nuclear changes In the mean time the central mass of medullary cells loses its solid aspect The medullary cells cover the internal side of the cortex as a flat layer around a central cavity the ovarian sac

Testicular organogenesis Testicular organogenesis on the contrary is characterized by the proliferation of the medullary cells These differentiate testicular ampullae which first appear as outgrowths of the medullary mass Primordial germ cells become included in these ampullae while the cortex disappears being replaced by a thin layer of connective tissue around the testis, the albuginea

Some deviations from this general scheme should be mentioned In frogs, *the above-summarized proc*ess is seen only in animals living in cold countries or at high altitude In most frogs from temperate *climates the process of sexual differentiation is* different During the months before metamorphosis all individuals differentiate ovaries, whatever their *genetic sex A sex reversal of the* gonad occurs later in 50% of the individuals The ovarian sac (medulla) proliferates and differentiates testicular *ampullae*, which become inhabited by the smallest of the primordial germ cells The cortex and the large ovocytes degenerate Finally a true testis evolves after the ovarian phase of the gonad Animal species or strains which show such a transitory feminine phase of the male gonads are known as indifferent strains Several amphibians, teleosts, and cyclostomes also develop indifferent strains

Another particular case is noteworthy In toads, the anterior part of the gonadal anlage is almost entirely a cortical primordium with no medulla This primordium persists as a rudimentary organ,

Bidder's organ above the functional gonad In adult animals Bidder's organ may develop into a functional ovary when the actual gonads are surgically removed Males deprived of their testes undergo a slow feminization and may lay eggs (K Ponse)

Gonad organogenesis. Organogenesis of the gonad in reptiles, birds and mammals exhibits another pattern which makes the detailed interpretation more difficult

When the indifferent stage of the gonad appears, it also consists of a cortex which is a germinal epithelium derived from the coelomic epithelium, and a medulla comprised of cords of cells included in the general connective stroma of the organ The primordial germ cells become distributed in both the medullary cords and the germinal epithelium

Testes During testicular differentiation, the medullary sex cords increase in size and directly give rise to the future testicular tubules, in which spermatozoa will later mature Between these seminal tubules, glandular cells appear, the interstitial cells which secrete the male sex hormone The cortex disappears and leaves a peripheral layer of connective tissue, the albuginea of the testis

Ovaries Ovarian differentiation is first characterized by a prolonged proliferation of the cortex Secondary sex cords are progressively pushed into the underlying connective tissue while the medullary sex cords retrogress The secondary sex cords will later divide into small clumps, the primordial follicles, each one comprising one primitive germ cell surrounded by follicular cells Each follicle becomes the functional unit of the adult ovary

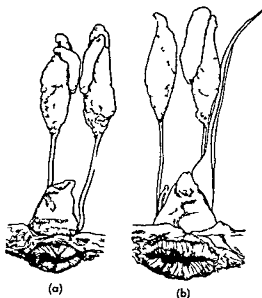


Fig 5 Reproductive system of normal chick embryos, shortly before hatching (a) Male (two testes present) (b) Female (development of sole left ovary and left oviduct) (After B H Willier, in E Allen, ed, Sex and Internal Secretions, 2d ed, Williams and Wilkins 1939)

Gonadal differentiation is well established in reptiles or birds before hatching or in placental mammals at birth In the marsupials, such as the opossum intrauterine pregnancy is very short and is followed by a period of development in the mar

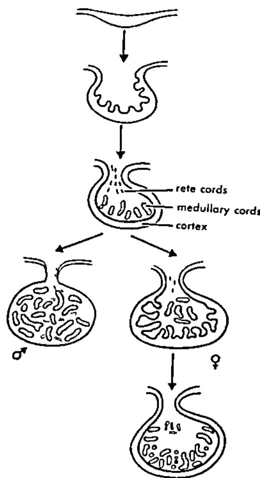


Fig 4 Schematic presentation of the differentiation of the gonad in higher vertebrates

assumed that the medullary cords appear as outgrowths from the inner surface of the germinal epithelium Later these cords become connected with some mesonephric urinary tubules by cords of cells which constitute the rete cords In the adult testis the rete testis constitutes an intratesticular net of canals in the epididymis it is connected by former mesonephric tubules with the vas deferens It has been suggested by E Witschi that the first origin of the medullary cords of the indifferent gonad and of the rete should be seen in cords of cells growing from mesonephric glomeruli towards the germinal epithelium If such an interpretation were substantiated a more common scheme would apply to gonadal organogenesis in both lower and higher vertebrates The medulla the prospective testicular component would then originate from the mesonephric blastema, and the cortex or the prospective ovarian component would originate from the coelomic epithelium

Birds The particular case of the right gonad of birds should be mentioned In a great number of birds only one ovary becomes functional in the adult In the female chick embryo for instance di-

ymmetry of the gonads is conspicuous at an early age because the left anlage is much larger than

small testis if the left ovary is removed

Gonadal abnormalities Important abnormalities include absence of the gonad hermaphroditism, and sex reversal

Absence of the gonad The entire gonad may fail to develop or may retrogress at very early stages leaving only some indistinct remnants which are impossible to recognize as testes or as ovaries This condition, called gonadal agenesis or gonadal dysgenesis, is known among humans as Turner's syndrome and has also been observed among animals such as pigeons.

In 1956 C. Houillon was able to produce gonadal agenesis in the newt *Pleurodeles waltlii* by interfering with the organogenesis of the mesonephros This was produced by stopping the early growth of the primitive ureter before it reached the mesonephric blastema In the absence of the inductive influence of this duct the differentiation of the mesonephros may be impaired When no mesonephros develops the gonad fails to differentiate because the established germinal epithelium suffers retrogression It remains to be ascertained whether such relations between the mesonephros and the gonads obtain for other vertebrates

Hermaphroditism The presence of male and female gonadal tissue in the same individual is not normal among vertebrates except in some species of fishes. It can occur as an abnormal condition more or less frequently among other vertebrates It is rather frequent in some frogs and toads and is considered to be the result of an incomplete dominance of either the cortex or the medulla of the indifferent gonad It is also frequent in the pig but rather rare in humans

Sex reversal of the gonad Sex reversal of the gonad, which results from the differentiation of the gonad in a direction opposite to the genetic sex may be produced experimentally According to Witschi's scheme the cortex and medulla of the undifferentiated gonad constitute a pair of antagonistic inductors which compete by producing one or more inductive substances (cortexin and medullarin) The final sex of the gonad results from the dominance of either the cortex (female) or the medulla (male) Under experimental conditions whatever the genetic sex of the animal any condition which depresses the cortex or gives prevalence to the medulla results in testicular differentiation depression of the medulla or prevalence given to the cortex results in an ovary

Competition between the sex inductors may be observed if two developing larvae are united in parabiosis in such a way that exchange of blood occurs or if the gonadal anlage from one embryo is grafted to another embryo

Sex reversal namely masculinization of a female urodele larva under the influence of the testis of another individual or partial feminization of a male chick by a grafted ovary was experimentally obtained by E Wolff

Sex reversal was also obtained in experiments in which adult sex hormones were administered to developing embryos Results are quite variable according to the animal species and the hormone used Examples of such experiments are included in the table

Effect of administered sex hormones on developing gonads

Animal	Estradiol (on males)	Testosterone (on females)
Fish		
Lebistes	Feminization	Masculinization
Amphibians		
Frogs, various species	Feminization (low dosage) Masculinization (high dosage) paradox effect	Masculinization
Alytes	Feminization	No effect
Xenopus	Complete feminization	No effect
Amblystoma	Complete feminization	Feminization paradox effect
Pleurodeles	Complete feminization	Feminization paradox effect
Birds		
Chick	Feminization	Slight masculinization
Duck	Slight feminization	No effect
Marsupials		
Opossum	Feminization	No effect
Placental mammals		
Several species	No effect	No effect

Genital tract development. The sex ducts become sexually specialized some time after the sexual differentiation of the sex glands The male or female conditions develop from an indifferent condition which is identical in both sexes in early stages

Indifferent stage The gonads are already recognizable as ovaries or testes and are located on the anterior part of the mesonephros The mesonephric or Wolffian duct is the ureter, and opens posteriorly into the cloaca in lower vertebrates or into the urogenital sinus in mammals Another duct the oviduct or Mullerian duct parallels the mesonephric duct

The oviduct arises from a funnel which opens into the coelomic cavity The blind end of this primordium proliferates and extends progressively caudally It was assumed at one time that the Mullerian duct was derived from the Wolffian duct by longitudinal division Such an assumption seems to be valid only for selachians

In selachians and urodeles the funnel from which the Mullerian duct originates corresponds to a pronephric nephrostome the coelomic opening of the primitive urinary tubules Because the pronephros is located near the neck of the larva and because the ostium of the oviduct retains this position the oviducts open into an anterior part of the body cavity (see Fig 8)

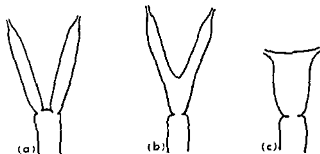


Fig 6 Different types of uterus, resulting from a more or less complete fusion of the Mullerian ducts. (a) Uterus duplex (rat) (b) Uterus bicornis (mare) (c) Uterus simplex (human)

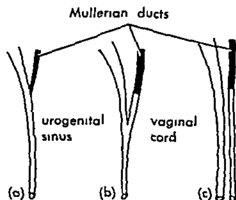


Fig 7 (a-c) Schematic interpretation of vaginal organogenesis

In birds and mammals the origin of the oviduct from pronephric remnants is not clear, but it is obvious that the oviduct develops in the region of the nephric field The early funnel is located on the top of the mesonephros and the ostia tubae open above the ovaries

Differentiation of female genital tract The Mullerian ducts differentiate into the female ducts Depending upon the animal species, either a simple secretory oviduct (amphibians) or a more complicated structure develops from this simple unicellular layered duct It is divided into several specialized sections involved in the secretion of albumen or shell as in selachians reptiles and birds In those birds in which the female has only one functional ovary, only one oviduct develops (see Fig 5)

In mammals, the Mullerian ducts give rise to the oviduct, or tube, and the uterus Usually the embryonic Mullerian ducts fuse posteriorly, but the extent of this fusion is variable In man and monkeys one single uterus is formed from the fused part of the ducts, the tubes correspond to the non fused part In rodents the two Mullerian ducts fuse only in the upper vagina, and form two uterine horns and two tubes In ruminants, an intermediary condition is realized in which the uterus is composed of an inferior stem and two horns which largely communicate

The posterior part of the mammalian female genital tract is constituted by the vagina The embryology of this organ displays great variability from one animal species to the other, and this makes a accurate interpretation difficult

In the undifferentiated stage, the Mullerian duct terminate blindly at the wall of the urogenital sinus (or embryonic urethra), between the tw Wolffian ducts which open into the urogenital sinus They may retain this primitive connection, as is the case in the rabbit (Fig 7a), and the urogenital sinus then becomes a urethrovaginal duct

In other animals a cord of cells, the vaginal cord may detach from the dorsal wall of the urogenital sinus and grow progressively caudally below th end of the Mullerian parts (Fig 7b) In the rat o the mouse this sinusary vagina finally opens independently of the definitive female urethra in a vaginal opening (Fig 7c) In other animals, such a the mare, the vaginal cord remains connected wit the posterior part of the urogenital sinus. The vestibulum then remains as a common opening to both the vagina and urethra

Comparative embryology of the vagina helps in understanding abnormalities of the human genital tract Absence of the vagina and opening of the female ducts into the urethra may result from the preservation of the primitive connections

The Wolffian or mesonephric duct of the female embryo is never incorporated into the genital tract In animals which keep their mesonephros a a functional kidney in adulthood (selachians and amphibians) the mesonephric duct remains as the ureter (Fig 8) In female birds and mammals, the mesonephric duct disappears as well as the mesonephros, leaving only some minor vestiges.

Differentiation of male genital tract The male sex ducts are derived from the mesonephric or Wolffian ducts In selachians and amphibians the *mesonephros remains the functional* kidney, and the mesonephric ducts function as pathways for urine as well as for sperm Even in such animals however the anterior part of the kidney often becomes specialized as the sexual part The posterior part of the kidney then produces urine, and several excretory tubes may bring this urine directly to the cloaca In such cases the Wolffian duct of the male is only a genital canal as in the newt *Triturus cristatus* (Fig 8)

The Wolffian ducts of birds differentiate into an undulated vas deferens whereas in mammals the ducts differentiate into the epididymis at one end and the seminal vesicles at the other (Fig 10)

The Mullerian duct has no function in males, and as a rule it disappears In some amphibians, such as newts and toads, the oviducts persist in a rudimentary condition (Fig 8), they may be activated under appropriate hormonal stimulation in adult males

In male mammals the urogenital sinus becomes the definitive male urethra Several accessory glands, such as the prostatic glands, bud from it, and display great variations from one animal species to another, but all open into the urethra

Copulatory organs. Copulatory organs are well developed in mammals and reptiles In selachians

they are often a specialized part of the fins the cla pers. As a rule no copulatory organ is present in amphibians and birds only a few male birds such as the duck possess a penis

In mammals the copulatory organ develops from an undifferentiated genital tubercle which is iden tical in both the male and female embryos and lies abo e the opening of the urogenital sinus The male penis encloses the penile urethra and increases in size, whereas the homologous female clitoris re mains more l ke the primitive tubercle

Hormonal control The genital tract differenti ates after sexual differentiation of the sex glands and time relationships support the view that sex ual specialization of the genital tract is controlled

the gen tal tract becomes identical whatever the

lebrate species does not appear This identical hor moneless aspect of the body is known as the neu tral form.

In the newt *Triton cristatus* J de Beaumont (1933) noticed that the neutral form is more or less intermediate between the male and female con dition, but that there is a definite trend toward ac qui t on of masculine features such as a dorsal crest, cloacal glands and fusion of the urinary col lecting tubules (Fig 8) *During normal sexual de velopment of females the ovaries inhibit such* masculine characters

In the duck embryo castrated by a beam of x rays E Wolff (1951) noticed an even more defi nite masc iline differentiation of such sex charac ters as the

sexual di

such char

tion the neutral aspect of the genital tract is also characterized by the presence of oviducts These structures are normally absent in males where they *are inhibited by the embryonic testicular hormone* (Fig 9)

nine This means that no embryonic gonad is nec essary to produce the feminine sexual structures In males the testes prevent persistence and devel opment of the female structures (tube and uterus) and impose masculinity on the whole genital ap *paratus* (Fig 10)

In agreement with such an experimental obser vation human beings in which gonads failed to develop (Turner s syndrome) display complete feminine features although secondary sex charac *ters remain infantile because of the absence of sex* glands

It appears that in some vertebrates the neutral type is predominantly masculine (urodeles and

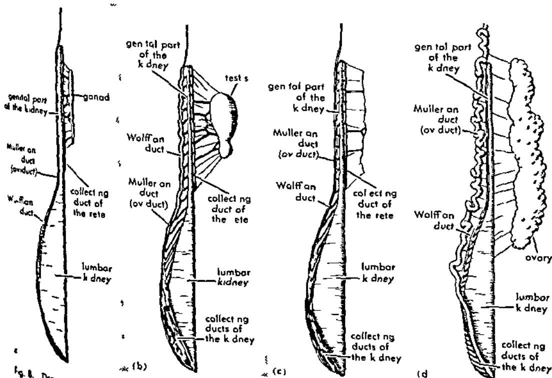


Fig 8. Draw ngs of the right half of the reproductive tract of the newt Tr ton (*Triturus*) *cr status* (a) Undifferentiated cond t on (b) D fferent ated male (c) Neutral cond tion of a castrated an mal (d) The d fferent ated female (*From J de Beaumont W lhelm Roux Arch Entwicklungsmech Organ* 129 120 1933)

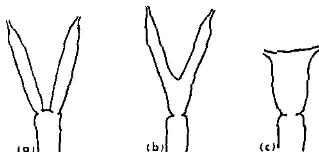


Fig 6 Different types of uterus, resulting from a more or less complete fusion of the *Mullerian ducts* (a) Uterus duplex (rat) (b) Uterus bicornis (mare) (c) Uterus simplex (human)

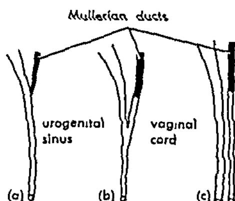


Fig 7 (a c) Schematic interpretation of vaginal organogenesis

In birds and mammals the origin of the oviduct from pronephric remnants is not clear, but it is obvious that the oviduct develops in the region of the nephric field The early funnel is located on the top of the mesonephros and the ostia tubae open above the ovaries

Differentiation of female genital tract The Mullerian ducts differentiate into the female ducts Depending upon the animal species either a simple secretory oviduct (amphibians) or a more complicated structure develops from this simple unicellular layered duct It is divided into several specialized sections involved in the secretion of albumen or shell as in selachians reptiles, and birds In those birds in which the female has only one functional ovary, only one oviduct develops (see Fig 5)

In mammals, the Mullerian ducts give rise to the oviduct, or tube, and the uterus Usually the embryonic Mullerian ducts fuse posteriorly, but the extent of this fusion is variable In man and monkeys one single uterus is formed from the fused part of the ducts, the tubes correspond to the non fused part In rodents, the two Mullerian ducts fuse only in the upper vagina and form two uterine horns and two tubes In ruminants, an intermediary condition is realized in which the uterus is composed of an inferior stem and two horns which largely communicate

The posterior part of the mammalian female genital tract is constituted by the vagina The embryology of this organ displays great variability from one animal species to the other, and this makes ar accurate interpretation difficult

In the undifferentiated stage, the Mullerian dı terminate blindly at the wall of the urogen sinus (or embryonic urethra), between the ı Wolffian ducts which open into the urogenital nus They may retain this primitive connection, is the case in the rabbit (Fig 7*a*), and the urog ital sinus then becomes a urethrovaginal duct.

In other animals a cord of cells, the vaginal co may detach from the dorsal wall of the urogen: sinus and grow progressively caudally below : end of the Mullerian parts (Fig 7*b*) In the rat the mouse this sinusary vagina finally opens in pendently of the definitive female urethra in a v inal opening (Fig 7c). In other animals, such the mare, the vaginal cord remains connected w the posterior part of the urogenital sinus The v tibulum then remains as a common opening to b the vagina and urethra

Comparative embryology of the vagina helps understanding abnormalities of the human gen tract Absence of the vagina and opening of the male ducts into the urethra may result from ı preservation of the primitive connections

The Wolffian or mesonephric duct of the male embryo is never incorporated into the gen tract In animals which keep their mesonephros a functional kidney in adulthood (selachians a amphibians) the mesonephric duct remains as t ureter (Fig 8) In female birds and mammals t mesonephric duct disappears as well as the me nephros, leaving only some minor vestiges

Differentiation of male genital tract The ma sex ducts are derived from the mesonephric Wolffian ducts In selachians and amphibians t mesonephros remains the functional kidney, a the mesonephric ducts function as pathways f urine as well as for sperm Even in such anima however, the anterior part of the kidney often b comes specialized as the sexual part The post rior part of the kidney then produces urine, ar several excretory tubes may bring this urine d rectly to the cloaca In such cases the Wolffia duct of the male is only a genital canal as in th newt *Triturus cristatus* (Fig 8)

The Wolffian ducts of birds differentiate into a undulated vas deferens whereas in mammals th ducts differentiate into the epididymis at one en and the seminal vesicles at the other (Fig 10)

The Mullerian duct has no function in males an as a rule it disappears In some amphibians, suc as newts and toads, the oviducts persist in a rud mentary condition (Fig 8), they may be activate under appropriate hormonal stimulation in adul males

In male mammals the urogenital sinus become the definitive male urethra Several accessor glands, such as the prostatic glands, bud from i and display great variations from one animal spe cies to another, but all open into the urethra

Copulatory organs. Copulatory organs are wel developed in mammals and reptiles In selachian

n teleosts The anterior portion ends blindly but
most cases continuations of the folds at the poste
... and form an oviduct which opens directly to

Amphibians Although amphibian ovaries are saccular structures, ripe ova are liberated *into the* body cavity through their external walls The shape of the ovaries varies with the shape of the body They are long and *narrow* in caecilians, elongated to a lesser degree in salamanders and short and more compact in frogs and toads Fat bodies are associated with amphibian ovaries They serve for the storage of nutriment and undergo profound changes *during the* year *A peculiar structure in* the male toad called Bidder's organ may under certain conditions develop into a true ovary

Reptiles The saccular ovaries of snakes and lizards are similar to those of amphibians and are in contrast to the solid ovaries of turtles and crocodilians. In snakes and lizards they are elongated but not symmetrically disposed Only the yolk of reptilian eggs is *formed in the ovaries,* and this represents the true ovum The size of the egg is in proportion to that of the animal In certain ovoviviparous snakes and lizards corpora lutea form from ruptured follicles after ovulation These probably secrete a *hormone necessary for maintenance* of pregnancy

Birds Although both ovaries are present during embryonic development in most birds (except many birds of prey) the right ovary degenerates and only the left is *functional* A mature ovum escapes from the ovarian follicle through a preformed nonvascular area on the surface which ruptures Increase in the number of hours of daylight stimulates ovarian activity in many birds

liberate their ova into the body cavity In such animals as the rabbit cat and ferret ovulation will not occur unless the *animal* copulates Following ovulation certain cells of the follicle undergo a transformation and the entire structure becomes a more or less solid body the corpus *luteum* If pregnancy does not occur the corpus luteum persists only for a short *time If* pregnancy does ensue the corpus luteum persists throughout pregnancy In either case *it* ultimately degenerates The corpus luteum is of primary importance as an endocrine gland secreting a hormone called progesterone



Fig 11 Section of portion of cortex of rat ovary, showing ovarian follicles in various stages of development (From C K Weichert Anatomy of the Chordates 2d ed McGraw Hill 1958)

Oviducts Oviducts except in teleosts and a few other fishes are modifications of Mullerian ducts formed early during embryonic development Although Mullerian ducts also form in the male they

a uterus

Cyclostomes Oviducts are lacking in cyclostomes Ova pass from the coelom through genital pores and out a urogenital papilla

Fishes Much diversity exists in regard to the oviducts of fishes In some teleosts and a few others eggs escape from the body cavity through modified *abdominal pores* In elasmobranchs the two Mullerian ducts may fuse at their anterior ends so that only a single aperture the ostium tubae opens into the body cavity An enlarged shell gland is present in each oviduct Some elasmobranchs lay eggs *but others are ovoviviparous* The uteri open separately into the cloaca The oviducts of most teleosts are short and are continuous with the cavities of the saccular ovaries It is doubtful whether they are true Mullerian ducts because they are formed in a different manner A cloaca is lacking in teleosts and the oviducts open independently to the outside The two oviducts often fuse at their *lower ends and have a common opening Most* teleosts are oviparous but many are ovoviviparous The young may even develop within the cavities of the ovaries but intrauterine development is more common The size of the oviducts fluctuates markedly *with the seasons They are largest during* the breeding period

Amphibians Oviducts *in amphibians* are paired elongated tubes each with an ostium situated well forward in the body cavity The posterior end of *each oviduct is enlarged slightly to form* a uterus which opens into the cloaca In some toads the two oviducts unite before entering the cloaca by a common orifice Marked fluctuation in size of the oviducts is apparent at different seasons (Fig 12) The *glandular lining secretes* a clear gelatinous substance (jelly) which is deposited about each

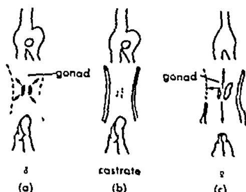


Fig 9 Scheme of evolution of sex characters in the duck embryo (a) male, (b) castrate, (c) female First row, syrinx, second row, Mullerian ducts, third row, genital tubercle Arrows symbolize inhibiting gonadal actions (From E Wolff, *Compt rend*, 229 428, 1949)

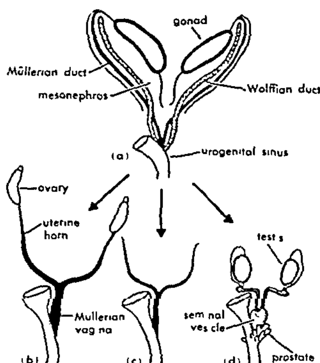


Fig 10 Sexual differentiation of the genital tract of the rabbit fetus *(a)* undifferentiated condition, *(b)* female *(c)* castrate, *(d)* male (From *A Jost, Mem Soc Endocrinol*, in press)

birds) and in others feminine (mammals) Because the homozygous sex is masculine in most of the urodeles and birds and feminine in mammals some correlation is suggested between the characters of the neutral sex type and the type of genetic sex determination, but no definitive statement is yet possible

Extraneously administered sex hormones may in some instances completely reverse the sexual differentiation of the genital tract in amphibian larvae Genetic males of the newt *Pleurodeles waltlu*, for instance, when adequately treated with estradiol, may develop a complete female sex apparatus, behave as females, and lay eggs which may be fertilized by another normal male.

Such a complete sex reversion has not yet been produced in higher vertebrates In mammals, for instance, androgenic sex hormones administered to the pregnant female may to a large extent masculinize the genital tract of the female young, but such animals keep their ovaries, tubes, and uteri in addition to masculine sex structures, they then become intersexes and definitely abnormal. The extraneously administered sex hormones do not exactly reproduce the effects of the fetal hormones. [A.J.]

COMPARATIVE ANATOMY

Egg cells, or ova and sperm cells, or spermatozoa, are formed in the primary reproductive organs which are collectively known as gonads. Those of the male are called testes, those of the female are ovaries. Besides giving rise to reproductive cells both ovaries and testes give off endocrine secretions, or sex hormones, which have a profound effect in the development, maintenance, and function of the rest of the reproductive system In both sexes the structures used to transport the reproductive cells and which serve to bring the cells produced by the two sexes together are known as the accessory sex organs. Secondary sex characters are those which serve to distinguish the sexes but are not directly concerned with sex. *See* REPRODUCTION ANIMAL.

Ovaries. A typical ovary is a solid, irregularly shaped structure indistinctly separated into an inner medulla and an outer cortex. The cortex contains numbers of ovarian follicles in various stages of development An egg cell lies within each follicle After growing to a certain extent some follicles push to the surface of the ovary Such follicles either rupture liberating the ovum into the coelom (ovulation) or the follicle and its contents degenerate (atresia) The size of the ovarian follicle depends mostly upon the volume of the ovum characteristic of the species. *See* OVUM

In certain fishes and in amphibians, snakes, and lizards, the ovaries are hollow saccular structures. The cavity within the teleost fish ovary is actually a closed-off portion of the body cavity into which ripe ova are shed This is not true of the saccular ovaries of other forms.

Cyclostomes The adult female lamprey has a single ovary representing a fusion of two, which courses the length of the body cavity suspended from the middorsal body wall by a single mesovarium At the height of the breeding season it fills the greater part of the body cavity The hagfish is hermaphroditic the anterior part of the single gonad is ovarian and the posterior part testicular Usually only one or the other region matures.

Fishes The ovaries of most fishes are paired although in some cases they have fused into a single organ The large eggs of elasmobranchs are discharged from the anteriorly located ovaries directly into the body cavity In ovoviviparous elasmobranchs, following ovulation the ovarian follicles

nor end of the body cavity In most other fishes they are elongated and often lobulated Ducts serve to transport spermatozoa to the outside

Amphibians The testes of amphibians vary greatly in shape. In caecilians, each is elongated and resembles a string of beads The enlargements consist of masses of seminiferous ampullae connected by a longitudinal collecting duct In salamanders the testes are shorter and irregular in

gonads of male as well as ...

Reptiles Reptilian testes are usually round or oval in shape. Seminiferous tubules are long and convoluted In snakes and lizards one testis usually lies farther forward in the body cavity than the other

Birds The round or oval shape of the bird's testis characteristic In a few birds, such as the domestic fowl, the testes function throughout the year most birds are seasonal breeders Increase in number of hours of daylight stimulates spermatogenesis in certain birds and hence brings about testicular enlargement

Mammals In all mammals except monotremes oval shaped testes move from their place of origin to the pelvic region where they may remain permanently, or they may descend farther into a pouchlike scrotum In many seasonal breeders the testes are located in the scrotum only during the breeding period. The scrotum serves as a tempera

...anterior to the penis but in others it is posterior to that organ. In several mammals a relation between the number of hours of daylight and testicular activity has been demonstrated

Male ducts The ducts which in most vertebrates serve to transport spermatozoa to the outside of the body are the archinephric ducts or the Wolffian

...undergo changes in size in seasonal breeders (Fig 14) Reproductive ducts are lacking in cyclostomes. See KIDNEY

Fishes A variety of conditions is encountered in male fishes. In elasmobranchs the ductus deferens courses along the ventral side of the opisthonephros. In older individuals it is highly convoluted The posterior end is dilated, forming a seminal vesicle. The ducts of the two sides enter the cloaca through a common urogenital sinus Connection of testes and archinephric duct is by means of modified kidney tubules at the anterior end of the opisthonephros

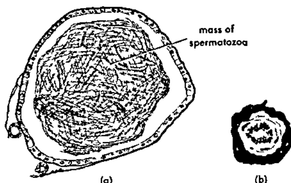


Fig 14 Cross section of ductus deferens of male salamander, *Eurycea bislineata* (a) From specimen captured just prior to breeding season (b) From specimen obtained several weeks after breeding season (From C K Weichert, *Anatomy of the Chordates, 2d ed*, McGraw Hill, 1958)

In most other fishes the kidney ducts serve only for the passage of urinary wastes The sperm duct, which is not a true ductus deferens because it is formed in a different manner, usually is entirely independent of the kidney duct, although the two may have a common opening to the outside

Amphibians The arrangement of the male ducts in amphibians rather closely resembles that of elasmobranch fishes In salamanders the ductus deferens courses outside the lateral border of the kidney Small tubules from the testis join Bidder's canal a duct lying medial to the kidney This in turns connects to the ductus deferens by means of modified kidney tubules In frogs and toads both Bidder's canal and the ductus deferens lie within the opisthonephric kidney The ductus deferens (archinephric duct) opens into the cloaca

Reptiles When the embryonic mesonephros degenerates in male reptiles its duct (archinephric or Wolffian duct) persists as the male reproductive duct The portion near the testis connects with that organ by means of a few persistent mesonephric tubules This end becomes highly convoluted, forming the epididymis The remainder is the ductus deferens which in snakes and lizards joins the ureter of the metanephros before entering the cloaca In turtles and crocodilians these ducts open at the proximal end of a groove which carries spermatozoa to the free end of the penis

Birds The reproductive ducts of birds are essentially similar to those of reptiles but open independently into the cloaca In the few birds possessing a penis a groove on the upper surface carries spermatozoa to the apex

Mammals A few persistent mesonephric tubules connect the seminiferous tubules of the testes to a compactly coiled epididymis which is continuous with the ductus deferens In those having scrotal testes the ductus deferens enters the body cavity, crosses in front of the ureter loops over that structure and then courses posteriorly for a short distance before joining the urethra (Fig 15) In sev-

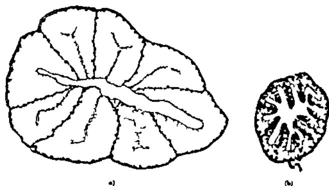


Fig 12 Cross section of oviduct of salamander, *Eurycea bislineata* (a) During breeding season (b) After breeding season (From C K Weichert Anatomy of the Chordates, 2d ed, McGraw Hill, 1958)

ovum as it passes down the oviduct External fertilization is the general rule in frogs and toads but in salamanders, with few exceptions, internal fertilization takes place No copulatory organs are present in either case A diverticulum of the salamander's cloaca serves as a storage place for spermatozoa The males deposit little packets of spermatozoa which are taken into the cloaca of the female by muscular movements of the cloacal lips Internal fertilization occurs in caecilians

Reptiles The paired oviducts of reptiles open into the body cavity through large slitlike ostia Each oviduct is differentiated into regions which mediate different functions in forming the envelopes deposited about the ova prior to laying The eggshell of oviparous reptiles is formed in the uterus Fertilization is always internal in reptiles Most are oviparous but many snakes and lizards are ovoviviparous

Birds The right oviduct usually degenerates in birds and only the left one is functional Birds of prey are exceptions The egg enters the oviduct through the ostium and passes through a glandular region an isthmus and a uterus Albumen is deposited about the ovum in the glandular region, inner and outer shell membranes and more albumen are laid down in the isthmus the hard limy shell is formed in the uterus Fertilization is internal in birds all of which are oviparous

Mammals Paired Mullerian ducts develop in all mammalian embryos Each differentiates into an anterior narrow Fallopian tube and a posterior expanded uterus In all except monotremes the uterus leads to a terminal vagina which serves for the reception of the penis of the male during copulation. Marsupials retain the primitive paired condition the two vaginae opening into a common urogenital sinus Placental mammals have a single vagina which represents a fusion of two The uterine portions fuse to varying degrees, resulting in different types of uteri (Fig 13) The simplex type found in apes and man represents the greatest degree of fusion A small erectile organ, the clitoris comparable to the penis of the male, lies ventral to the vaginal orifice It differs from the penis, however, in that it has no connection with the urethra except in a few forms such as rats and mice

Certain glands as well as remnants of the degenerated mesonephric kidney and Wolffian duct are associated with the female reproductive system The glands of Bartholin correspond to Cowper's glands of the male, secreting a clear viscid fluid under sexual excitement

Testes. The typical testis is a compact organ which varies greatly in shape in different groups of vertebrates In all but a few primitive forms each testis is composed of numbers of seminiferous ampullae or seminiferous tubules which connect by means of ducts to the outside. Spermatozoa are formed within the tubules or ampullae In addition to sperm production the testes of vertebrates are endocrine organs which secrete a sex hormone testosterone The actual tissue which secretes testosterone has not been definitely determined in the lower classes In mammals, groups of interstitial cells lying among the seminiferous tubules are undoubtedly the endocrine elements

In seasonal breeders the size of the testes fluctuates with the seasons, being largest just before the breeding season After this period they shrink to only a fraction of their former size Each testis is suspended from the middorsal body wall by a membrane, the mesorchium

Cyclostomes The testis of the lamprey is a single organ representing fusion of two It is attached to the middorsal body wall by a single mesorchium Spermatozoa are discharged into the body cavity from which they escape through genital pores The hermaphroditic gonad of the hagfish has already been mentioned

Fishes In elasmobranchs, the testes are relatively small paired structures located at the ante

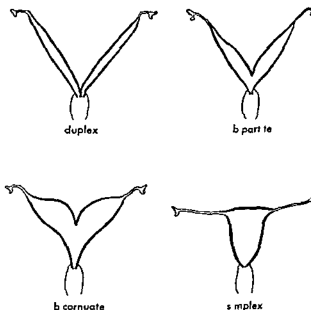


Fig 13 Diagram illustrating degrees of fusion of the uterine portions of the Mullerian ducts in four types of mammalian uteri (From C K Weichert Anatomy of the Chordates 2d ed, McGraw Hill 1958)

along the midline of the abdomen, usually in a hor izontal position In most forms it lies within a sheath from which it may be protruded and re tracted. In primates it is permanently exserted The distal end of the penis or glans, is covered by a fold of integument, the foreskin There are many variations in shape and structure of the glans

[C K W]

HISTOLOGY

Male reproductive system The male system includes the testes scrotum excretory ducts, aux iliary glands, and penis

Testes The testes have two main products sper

tein hormones produced by the anterior lobe of the hypophysis cerebri (pituitary gland) and by the cooling action of the scrotum The testes have two important structural elements the contorted seminiferous tubules and the interstitial cells which are situated between the tubules The contorted seminiferous tubules have two kinds of cells nu

genesis. The first phase of spermatogenesis called spermatocytogenesis, involves cell division and structural changes and yields spermatids which have one-half the number of chromosomes found in somatic cells, the early sex cells or spermatogonia, become spermatids by passing through two intermediate stages namely, primary spermatocytes and secondary spermatocytes In the second phase of spermatogenesis called spermiogenesis the spermatids do not divide, instead, they undergo maturation and become spermatozoa The intersti tial cells of the testis are seemingly modified con nective tissue cells They resemble epithelial cells and hence are correctly designated epithelioid cells. According to experimental observations they produce androgen, or male sex hormone, which chemically is a steroid compound probably testos terone. See SPERMATOGENESIS

Scrotum The scrotum a pouch which contains the testes and spermatic sac, consists of skin smooth muscle, and connective tissue The scrotal skin is rich in sweat glands hence it acts as a thermoregulator for the testes The human scrotal temperature is about 7C° lower than that of the abdominal cavity The smooth muscle is a "skin muscle," called the dartos tunic its relaxation lengthens the scrotum and promotes loss of heat whereas its contraction shortens the scrotum and reduces loss of heat See SWEAT GLAND THERMO-REGULATION

Excretory ducts The excretory ducts which con vey spermatozoa from the contorted seminiferous tubules to the urethra are either intratesticular (tubuli seminiferi recti and rete testis) or extra testicular (epididymis, ductus deferens and ejacu latory duct) Spermatozoa first pass from the contorted seminiferous tubules into the tubuli semi niferi recti or straight seminiferous tubules, and

testis Next the spermatozoa enter the epididymis a tubular structure consisting of head, body, and tail The head of the epididymis has 12 or more efferent ductules (ductuli efferentes testis), the epithelium of which bears cilia, or little hairlike projections, which assist in moving sperm into the coiled part of the epididymis The body and tail of the epididymis constitute the ductus epididymidis, a coiled tube which when uncoiled is about 15–20 ft long The ductus deferens (vas deferens), a part of the spermatic cord is a tube in which convolu tions are mostly in its upper part Its total length, when uncoiled is about 18 in Its lower end is di lated to form the ampulla ductus deferentis The ductus deferens terminates by opening into a short duct the ejaculatory duct, which in turn "pierces" the prostate gland and opens into the prostatic part of the urethra

Auxiliary glands The auxiliary glands include two seminal vesicles one prostate, and two bulbo urethral glands Their secretions together with spermatozoa and the small amount of secretion from the excretory ducts constitute the semen In man one ejection of semen called ejaculation, has a volume of 3–5 ml and each milliliter contains about 60 000 000 spermatozoa The seminal vesi cles, named in accordance with the erroneous belief that they are receptacles for spermatozoa (after death sperm may be found in the vesicles), are hollow glands which lie along the back wall of the prostate They are about 2 in in length and devel opmentally each gland is an outgrowth of the am pulla of the ductus deferens The right and left seminal vesicles open into the prostatic urethra via the ejaculatory ducts The prostate is a solid gland which surrounds the urethra (prostatic urethra) at its origin from the urinary bladder It is about 1–1¼ in in diameter and has several small ducts which open individually into the prostatic urethra The bulbourethral glands or glands of Cowper, are about ¼ in in diameter Each opens into the ure- thra at a site slightly below the prostate

In such species as the rat guinea pig and rhesus monkey the mixture of the secretions from the aux iliary glands coagulates as in the case of ejacu lated semen In the mating of rats and guinea pigs the coagulated semen produces the vaginal plug which temporarily partly occludes the vagina of the female partner

Penis The penis or male organ of copulation, consists of two roots attached to bone (crura pe nis) a body (corpus penis) and a conical tip (glans penis) The corpus has three cylinders of erectile tissue right and left cavernous bodies (corpora cavernosa penis) and the unpaired cav ernous or penile, portion of the urethra (corpus spongiosum penis) The erectile tissue is composed of a spongelike system of blood sinuses (spaces)

eral mammals the ductus deferens is enlarged at its posterior end A glandular seminal vesicle, which contributes to the seminal fluid, often connects to the ductus deferens near its junction with the urethra Seminal vesicles are absent in monotremes, marsupials, carnivores, and whales The urethra coming from the bladder extends the length of the penis Accessory glands associated with the urethra include the prostate gland, which contributes to the seminal fluid, Cowper's glands, which secrete a clear viscid fluid during sexual excitement, and the small mucus-secreting urethral glands A few vestigial remnants of the mesonephros may persist in males in close relation to the reproductive system Persistent portions of the Müllerian ducts are also often present

Copulatory organs. Although external fertilization occurs in many lower aquatic vertebrates, internal fertilization is the rule in terrestrial forms and even occurs in numerous aquatic species Copulatory organs are usually employed to deposit spermatozoa, suspended in seminal fluid, within the reproductive tract of the female *See* COPULATORY ORGAN

Fishes In those fishes having internal fertilization, the copulatory organs are modifications of the pelvic fins (elasmobranchs) or anal fins (tel[e]osts)

Amphibians Although internal fertilization o[c]curs in most salamanders, copulatory organs a[re] lacking By muscular action of the cloacal lips th[e] female is able to pick up packets of spermatozo[a] deposited by the males The eversible cloaca o[f] some caecilians may be used as a sort of copul[a]tory organ

Reptiles *Sphenodon* is the only reptile lackin[g] copulatory organs In others, two types of structur[es] are recognized Snakes and lizards employ paire[d] hemipenes which are saclike structures, devoid [of] erectile tissue, lying under the skin adjacent to th[e] cloaca They are not comparable to the single pen[is] of turtles and crocodilians which is basically sim[i]lar to that of mammals, and which becomes di[s]tended with blood during sexual excitement.

Birds Most birds copulate by cloacal appositio[n] A penis is present only in ducks, geese, swans, an[d] ostriches It is similar to the organ of turtle[s,] crocodilians and mammals

Mammals In monotremes the single penis lies o[n] the floor of the cloaca, from which it may b[e] everted In most mammals there is a tendency f[or] the penis to be directed forward It is situate[d]

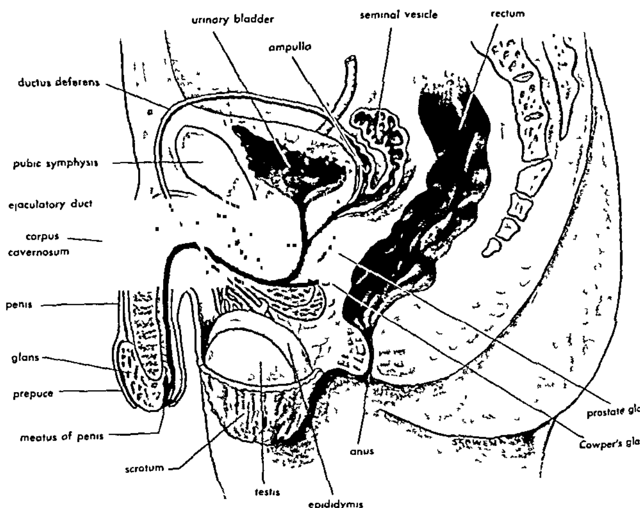


Fig 15 Urogenital system of the human male (From C K Weichert, Anatomy of the Chordates, 2d ed, McGraw Hill, 1958)

then divides differentiates grows, and develops into an individual either outside the mother (oviparity) or inside the mother (viviparity) The obvious advantage in bisexual reproduction lies in the fact that young produced from the mingled genes of two ancestral lines will be unlike either parent, representing different *combinations of ancestral traits* with great variations for the survival of the fittest. Reproduction is influenced or inhibited by environmental and nutritional factors as well as by all the other physiological activities but it is controlled predominantly by the endocrine system and mediated at least in part by the nervous system See FERTILIZATION

Breeding season The season when animals perform their reproductive functions is known as the *breeding season* This sexual periodicity is a general phenomenon common to plants and animals In general, sexual periodicity is predominant in lower vertebrates, but all birds and wild mammals are seasonal breeders As nutritional and environmental conditions improve, the reproductive season is not so restricted, cattle the domestic rabbit, and man all illustrate this fact, nevertheless their fertility is higher in spring than in winter Males of many mammalian species are capable of copulation at any time and *rarely experience a true sexual period* such as rut in the deer In mice and rats, testes do not descend to the scrotum until puberty and in certain other rodents they descend only during the breeding season Most vertebrates breed only in spring or summer *at the time when* food, temperature and light are optimal they are in the best physiological condition for reproduction and their offspring have the best chance of survival.

All environmental factors play a part but an increase of temperature for lower vertebrates and an

[illegible]

mer or in the *spring of* the following year The armadillo, deer goat and sheep breed *at the decrease* of daylight in the autumn and their offspring are born the next spring Transportation of sheep from the Northern to the Southern Hemi

[illegible]

[illegible] 50–100 times their weight in the inactive stage and the growth of accessory glands follows that of the testes. Although artificial increase of daylight brings wild rabbits and *ferrets into breed*ing condition in midwinter, permanent short day lighting does not entirely prevent breeding in the [illegible] Furthermore, artificial increase of daylight has so far failed to show any *definite effect* in advancing the breeding season of the domestic rabbit, guinea pig hedgehog or ground squirrel

Among fishes the duration of the breeding season varies considerably according to the group to which they belong and the location *where they live* *The ova of the elasmobranchs* are deposited singly or in pairs at varying intervals throughout a great part of the year but some species appear to have regular recurrent breeding seasons Most bony fish, producing millions of ova, breed only in the spring, and most of them *migrate to a suitable locality for* the deposition of their gametes In the frog and many other amphibians the ova are produced during the winter hibernation when the animals eat very little Similarly, the genital organs of salmon develop during migration *when the fish cease to feed* The spermatogenesis of the amphibian *Rana temporaria* is determined by the gonadotropic activity of the pituitary and by the sensitivity of the germinal epithelium to gonadotropin during the spring this activity and sensitivity are influenced *by temperature rather than light In the autumn* and winter their testes are not sensitive to gonadotropin under natural conditions Reptiles which hibernate usually begin to breed shortly after the beginning of the warm weather that terminates the *hibernating period Spring and summer are the sea*sons when most birds pair, build their nests and incubate their eggs The migration of some birds is invariably associated with an increase in the size of ovaries and testes

Estrus and menstrual cycles. The *cyclic* changes of reproductive activities in mammalian females are known as estrus or menstrual cycles Only at estrus (heat) will most mammalian females [illegible] ovaries grow at [illegible] Just *before or* [illegible] resulting in a higher output of estrogen a female hormone At the end of estrus the follicle ruptures and releases *the ovum (ovulation) The follicle is transformed* into the corpus luteum a glandular tissue which produces another female hormone progesterone If fertilization occurs the corpus luteum persists during all or most of the gestation period according to *the species Without mating or fertilization* the corpus luteum persists for a relatively short time and then degenerates The growth of the next crop of follicles begins again

Estrus Estrus in mammals can occur several times in one breeding *season*, *the mare*, ewe and rat come to estrus every 21 16, and 5 days respectively if breeding does not take place This condition is called polyestrus The bitch is monestrous, *she has only one heat or* estrus to the breeding season and if not served then she does *not come* into heat again for a prolonged interval 4–6 months according to different breeds The duration of estrus varies according to the species, such as 4–9 days in the mare, 13 *hours in the cow*, and 14 hours in the rat In the ferret, estrus is shown by predominant swelling of the vulva and lasts several weeks unless copulation takes place

which are situated between afferent arteries and efferent veins In erection these sinuses are filled with blood under pressure, thus leading to the enlargement and rigidity of the penis

Female reproductive system. The female system includes the ovaries, uterine tubes uterus, vagina, and external genitalia

Ovary The ovary (ovarium), or egg producer, is largely covered by peritoneum, it has follicles with egg cells (ova), a stroma of connective tissue, and, in maturity, corpora lutea

Young ovarian follicles either develop into mature follicles or degenerate (follicular atresia) A primary follicle is a spherical object consisting of a central ovum surrounded by a layer of follicular cells A growing follicle has several layers of follicular cells or granulosa cells (stratum granulosum), and a surrounding capsule of connective tissue the theca folliculi Older, growing follicles acquire vesicles by cavitation of the stratum granulosum and are called vesicular follicles (folliculi ovarici vesiculosi) A mature follicle, or Graafian follicle has a large cavity filled with follicular fluid It bulges the surface of the ovary, and its ovum is in the center of a hillock of granulosa cells the so called cumulus oophorus The ovarian follicles produce estrogen one of the female sex hormones

Rupture of the Graafian follicle is called ovulation In this process the superficial portion of the follicular wall on the surface of the ovary becomes thin and then ruptures The ovum enters the peritoneal cavity with the follicular fluid which flows over the surface of the ovary The recently shed ovum is surrounded by a crown of granulosa cells the corona radiata

Maturation of the ovum The process of the origin growth and formation of the ovum in its preparation for fertilization is known as oogenesis The young sex cell is an oogonium Growth produces the primary oocyte, only the beginning of the first maturation division occurs before ovulation After ovulation the primary oocyte divides to produce two unequal elements a secondary oocyte and a smaller body the first polar body Then the secondary oocyte divides to yield the mature ovum (rarely called ootid) and the second polar body The mature ovum has the haploid number of chromosomes, the two polar bodies eventually disintegrate *See* OOGENESIS

Corpus luteum The corpus luteum is a yellow endocrine body which originates at the site of a ruptured Graafian follicle and which produces progesterone one of the female sex hormones It originates by the metamorphosis of granulosa cells and thecal cells into lutein cells which are epithelioid The corpus luteum of ovulation persists about 14 days, whereas the corpus luteum of pregnancy lasts several months The degeneration of a corpus luteum produces a white, fibrous scar in the ovary, the corpus albicans

Fallopian tube The human uterine tube, or Fallopian tube, is a muscular structure about ½ in long, and has three regional subdivisions a funnel like upper part, the infundibulum (infundibulum tubae uterinae), a dilated part below the infundibulum, the ampulla, and a constricted part near the uterine junction, the isthmus The mouth of the tube, or ostium, communicates with the peritoneal cavity, and is guarded by a fringe, the fimbriae tubae The tube is the site where fertilization of the ovum occurs

Uterus The uterus in man has two portions, the corpus, or body, and cervix, or neck Between these parts is a transverse constriction, the isthmus uteri The uterus is covered in part by peritoneum It has an outer layer of smooth muscle, the myometrium and an inner layer of connective tissue and epithelium the endometrium The endometrium is a mucous membrane, the mucosa, and the thickness of the endometrium varies during the menstrual cycle

Vagina The musculomembranous organ situated between the cervix uteri and the external genitalia is the vagina During copulation, it ensheathes the penis Ventrally and dorsally, the anterior and posterior fornices, respectively, overlap the uterine cervix In youth the lower end of the vagina is constricted a bit by a membranous shelf of mucosa the hymen vaginae

External genitalia The external genitalia include the clitoris vestibule of the vulva (fossa vestibuli vaginae), paired lips of the vulva (labium majus pudendi and labium minus pudendi), and a pair of small vestibular glands (glandulae vestibulares minores) The clitoris is an erectile organ which is the homolog of the penis in the male It differs from the penis in that it does not include the urethra The large lips of the vulva are essentially folds of skin whereas the small lips are covered with a mucous membrane The vestibular glands secrete mucus and in cases of gonorrhea, may become inflamed and then form large cysts [L J W]

PHYSIOLOGY

Reproduction The physiological process by which a living being gives rise to another of its kind is considered one of the outstanding characteristics of plants and animals It is one of the two great drives of all animals self preservation and racial perpetuation In contrast to other physiological processes reproduction in vertebrates can be achieved only by two individuals the male and the female Each produces germ cells called gametes The male produces spermatozoa and the female ova which carry the physical materials (genes) in the chromosomes for the transmission of inherited characters No matter how discrepant the pairing gametes may be in size and in form they contribute the same number of chromosomes Although reproductive devices are quite different from one species to another all serve one end the bringing together of the spermatozoon and the ovum, each containing half the number of chromosomes of the parent After their union (fertilization) the fertilized ovum resumes the number of chromosomes of its race and

rtance in animal reproduction There are two rincipal components of gonadotropic hormones (1) the follicular stimulation hormone (FSH) which induces growth of follicles in the ovary of the female and spermatogenesis in the seminiferous tubules of the testes, (2) the luteinizing hormone (LH) which added to FSH causes the final growth and maturation of the large ovarian follicles leading to ovulation and the formation of corpora lutea LH also stimulates the activity of the interstitial cells of the testis to produce the male hormone, androgen, hence it is also called the interstitial cells stimulation hormone (ICSH)

There are other pituitary hormones which are closely related to reproduction, lactogens, or prolactin, galactin and mammotropin from the anterior lobe are essential to the secretion of milk by the crop glands of pigeons and by the mammary glands of mammals The luteotropic factor of lactogens maintains the corpus luteum and stimulates its production of progesterone Oxytocin se-

etabolized by the body tissue, especially by the ver and is destroyed or excreted in the less ac ne form of androsterone and other ketosteroids and related compounds The androgens stimulate the differentiation of male reproductive ducts at the embryonic stage and the growth of male accessory glands such as the seminal vesicles and the prostate gland. Secondary sex characters (physical

Androgens are also produced by organs other than the testes. Implantation of adrenal tissue in young cocks results in precocious development of male sex characters and behavior castration of young rats does not result in atrophy of accessory genital structures for many days unless there is simultaneous adrenalectomy Male accessory reproductive structures have been maintained in full functional activity in castrated specimens by grafted ovarian tissue *See* ANDROGEN

Estrogens The growing ovarian follicles also produce a steroid hormone known as estradiol one of the estrogens, which induces the enlargement and histological differentiation of the secondary sex organs, especially the uterus and the vagina The enlargement of follicles increases the output of estrogen, thereby causing estrus Estrogens are also produced by the testes of males The tissue with the highest known concentration of estrogens is the testis of the stallion, from which estrone has been isolated Another female hormone also a steroid, is progesterone, produced by the corpus luteum It has the effect of inducing progestational differentiation of the endometrium for the development and implantation of the early embryo and the maintenance of pregnancy The presence of corpora lutea, at least among the higher vertebrates, appears to be associated with viviparity Corpora lutea however, are present in the ovaries of certain elasmobranchs and reptiles, but they are absent in amphibians and their existence in birds is doubtful These luteal structures when present in lower vertebrates are not invariably associated with viviparity It is believed that another hormone, relaxin a water soluble ovarian extract which increases in the blood as pregnancy progresses is important to parturition It relaxes the pubic symphysis and the cervix, and inhibits uterine motility *See* ESTROGEN

General hormonal control It can be realized from the above account that the formation of gametes (spermatogenesis and ovagenesis) is controlled by anterior pituitary hormones The differentiation of male and female reproductive tracts is influenced and mating behavior and estrus cycles are controlled, by male or female hormones The occurrence of the breeding season is mainly dependent upon the activity of the anterior lobe of the pituitary which is influenced through the nervous system by external factors such as light and temperature The transportation of ova from the ovary to the Fallopian tube and their subsequent transportation development, and implantation in the uterus are controlled by a balanced ratio between estrogen and progesterone Furthermore it is known that estrogens androgens and progesterone all have the effect of inhibiting the production or the secretion or both of gonadotropic hormones thereby causing the cyclic changes of reproductive activity among different animals

Mammary glands are essential for the nursing of young Their growth production and secretion of milk and in fact the whole process of lactation are controlled by pituitary hormones as well as by estrogen and progesterone Other glands and physiological activities also influence lactation

Fish Endocrinology in relation to reproduction is better investigated in mammals than in other vertebrates however a few scattered features among vertebrates should be briefly mentioned In the smooth dogfish *Mustelus canis* ovulation requires at least several hours and perhaps days About 16–20 ova are released from the ovary, one at a time Hypophysectomy prevents ovulation or interrupts it once it is started but pituitary implants reinitiate it Seasonal variation in the histology of the pituitary correlated with the annual breeding period of teleost fish has been reported and cells responsible for the production of gonadotropins have been located in the middle glandular area of the anterior lobe of the pituitary Prema

In monestrous and seasonally polyestrous species the period of sexual quiescence between seasons is called anestrus. In sheep and deer this occurs in the summer, while in the ferret and mink it is restricted to the winter. Proestrus denotes a short period at the time of coming in heat. It is very marked in the bitch, the genital organs swell and become congested, and there is bleeding from the uterus. Following estrus, the ensuing period in which sexual activity declines is called metestrus. After metestrus a relatively longer period of diestrus occurs before proestrus recurs. By examining the cellular contents of vaginal smears C. Stockard and G. Papanicolaou demonstrated that all these phases can be determined in the guinea pig and rat. In other animals, cow, mare, and pig, for instance, the determination of estrus by smear is not accurate. Most females ovulate spontaneously, and shed their ova at the end of estrus: about 2–4 days before the end of estrus in the mare, 13 hours after the end of estrus in the cow, 10 hours after the beginning of estrus in the rat. Some animals such as the rabbit, ferret, and cat ovulate following the stimulation of copulation. If fertilization takes place, pregnancy occurs. If mating is unsuccessful, pseudopregnancy, a physiological condition very similar to pregnancy but which lasts a relatively short period, may occur in some species, particularly the dog and the rabbit. *See* ESTRUS.

Menstruation. The reproductive cycle of the female in the primate and human is well marked by menstruation, the period of blood flow. Menstruation does not correspond to estrus but occurs between the periods of ovulation at the time the corpus luteum declines precipitously. As a result of the collapse of the superficial capillaries in the endometrium, blood is extravasated in quantity and about two-thirds of the endometrium is desquamated or shed. After menstruation there is a period in which the uterus is replenished under the influence of estrogens by the growth of the epithelium and its capillaries. This is known as the stage of proliferation and is equivalent to proestrus and estrus in other mammals. It is accompanied by growth of the vaginal epithelium and some cornification but these changes are gradual and are not as well defined as in many other mammals. The Graafian follicle ruptures near the midcycle, about 15 days before the beginning of the next menstruation. Ovulation is spontaneous and is followed by the formation of the corpus luteum which causes great glandular growth of the uterus. This is called the progestational phase and is equivalent to diestrus in other mammals. In New World monkeys menstruation is confined to the appearance in the vaginal lavage of a few red blood cells. Tissue destruction and hemorrhage in the uterus of the elephant shrew, *Elephantulus*, of South Africa at the end of diestrus has been described, but the bleeding at proestrus in the bitch, at metestrus in the heifer 14 hours after ovulation, and at metestrus occasionally in the guinea pig is due to intensive congestion of blood vessels in the uterus or in the upper part of the vagina and is not true menstruation.

Mating. Mating, also called copulation or coitus, is the synchronized bodily activity of the two sexes which enables them to deposit their gametes in close contact. It is essential for successful fertilization because sperm and ovum have a very limited life span; as examples the fertilizing capacity of trout and salmon sperm lasts about 30 sec in fresh water, that of rabbit sperm about 30 hours in the female tract. The trout and frog ova become nonfertilizable soon after contact with water whereas the rabbit ova remain fertilizable for 6 8 h

tive vertebrate, the lamprey, some sort of mating takes place. The female is seized by the male who winds his tail around her and discharges his sperm over the ova as they extrude from her body. In the dogfish, which practices internal fertilization, contact between male and female is necessary for the claspers of the male to convey his sperm into the cloaca of the female. In the majority of the bony fish physical contact may be slight, yet the male follows the female closely or presses his body to hers and deposits his sperm over the ova either soon after they are laid or at the same time. Certain salamanders walk slowly ahead of the female and deposit spermatophores which the female straddles and secures with the lips of her cloaca. In other low vertebrates a rudimentary penis exists either as a single intracloacal organ (crocodile, turtle, and duck) or as a paired structure placed behind the cloaca (snake and lizard) and sperm is deposited into the female's cloaca at mating. In mammals the copulatory organs are well developed, varying a great deal in structure. Patterns of mating are different, the time required for mating varies from 3 hours (ferret) to a few seconds (sheep and rabbit), and the number of intromissions varies from 1 (bull) to 50 (hamster).

Mating in many birds and cold-blooded vertebrates is the culminating event of a more or less complicated series of love antics. It is quite probable that these sexual displays have a direct bearing upon ovulation and that in many species ovulation would not occur in the absence of such sexual experiences. Sexual excitement and violent physical exertions of both partners may also be essential for a fertile mating.

Endocrine function in reproduction. Glandular tissues which secrete certain substances (hormones) important to the growth, metabolism, or various physiological activities of other tissues or organs compose the endocrine system. The physiological activity of the reproductive system is predominantly controlled by the endocrine glands and their respective hormones.

Pituitary hormones. The pituitary gland situated underneath the hypothalamus of the brain produces in its anterior lobe several stimulating hormones. Among these are gonadotropic hormones, typical mucoproteins biochemically; they are of great im

complete thoracosympathectomy, absence of any nerve pathway to the ovaries, or cervical sympathectomy does not inhibit ovulation after coitus. Because stimulation of the brain, of the lumbo-sacral
sympa
induce
more than one nervous path and more than one mechanism for the initiation of ovulation must be involved. Furthermore, the rat, unlike the rabbit, ovulates spontaneously but a prolongation of the *life of the corpora lutea with subsequent pseudo-*pregnancy can be induced by sterile mating, mechanical stimulation of the cervix, or electrical stimulation of the brain. Pseudopregnancy in the rat also seems to be mediated by the pituitary through *nervous pathways. There are additional evidences* to show that the stimulus for LH release in a spontaneously ovulating animal is controlled by nervous mechanisms employing cholinergic and adrenergic components.

Neural humoral control. Other endocrine glands transplanted elsewhere in the body retain their essential normal functions but transplanted pituitary gland resumes efficient activity only when implanted under the temporal lobe of the brain or into the sella turcica of another hypophysectomized recipient. Thus the control of hypothalamus over pituitary functions via nervous or neurosecretory hormonal mechanism is obvious. It is highly probable that the anterior pituitary receives its messages humorally and transmits them in the same

Because essentially normal anterior gonadotropic and other functions were present when the only hypothalamic pituitary link was through connecting *blood* vessels, it has been emphasized that neurosecretions of hypothalamic origin carried as blood-borne hormones represented the essential neural humoral control mechanism of the anterior pituitary function. However, because there are other conflicting evidences, the possibility that there is more than one route of hypothalamic neural control of pituitary secretion is not entirely excluded. As for pathways from external stimuli to the reproductive activities of the organism as a whole through the intrinsic nerve tracts and nuclei of the hypothalamus to the anterior pituitary gland, controversies still exist and knowledge in this respect is not clear.

Lesions. It is well established that lesions in the basal tuber or median eminence induce ovarian atrophy in the cat, dog and rabbit. The role of the nervous system in establishing cyclic pituitary activity has also been emphasized. Localization of an erection center and of an ejaculation center in the hypothalamus has been reported. Appropriately placed hypothalamic lesions in the female guinea pig sometimes result in anestrus or prolonged estrus periods with sexual behavior *in keeping with the* gross changes of the cycle. In the male guinea pig, similar lesions induce sexual impotence, without genital regression. Hypothalamic production of oxytocin and the derivation of vasopressin from neurosecretory process have been postulated. As for mating behavior, it is assumed that advancing evolutionary status is accompanied by a progressive dominance of the nervous system and a corresponding reduction of endocrine control. To interpret the effect of sex hormones on mating behavior, F. Beach suggests that their activity may increase the excitability of the central excitatory mechanism.

Fertility and sterility. The ability or inability to produce offspring is termed fertility or sterility. Fertility and sterility occur in different grades among various species and among individuals of the same species. Absolute sterility is rare, but infertility of all degrees is very common, especially among higher vertebrates. The rate of reproduction in any species depends upon the average number of young born in each litter, the frequency of recurrence of breeding season, the duration of the reproductive period, and the age at which the animal starts to breed. The age as a general rule is earlier in small species than in large ones. In general, the number of young in a litter of mammals is inversely proportional to the size of the animal. For instance, a cow rarely produces twins, whereas the rat occasionally bears as many as 16 young. A theory of fertility proposed by H. Spencer is that individuation and genesis vary inversely, that is, the power to sustain individual life and the power to produce new individuals are inversely proportional. Where there is abundant food supply and a favorable environment and the necessary expenditure of energy is relatively slight, the cost of individuation is much reduced and the rate of genesis is correspondingly increased.

Factors controlling fertility. Hammond proposed that three factors control fertility:

1. The number of ova shed. In accordance with the genetic constitution of the species and the *nutritional status, the number of ova shed is* controlled by gonadotropic hormones through the pituitary gland, but influenced by external factors. Before puberty, at old age, and during pregnancy, pseudopregnancy, and lactation, practically no mammal ovulates. Although by administering gonadotropic hormones G. Pincus et al. demonstrated the possibility of increasing the number of ova shed in mammals, the actual number of young produced was rather low, probably as a result of other natural limitations.

2. The number of ova fertilized. This depends upon the number of spermatozoa produced by the male, and the morphological and physiological integrity of the gametes. It also depends upon the probability of meeting between gametes provided by the male and female in the lower vertebrates, the efficiency of sperm *transport to the site of*

ture spawning has been induced in several species of fish by injection of fresh pituitary extracts, and hypophysectomy in the killifish, *Fundulus*, is followed by regressive changes in the ovaries and testes Cells similar in origin and appearance to mammalian luteal cells have been observed following the ovulation and hypertrophy of cells of ruptured follicles in some viviparous fish In an oviparous fish, *Rhodeus*, these cells produce a progesteronelike hormone, oviducin, which is responsible for the development of the ovipositor Extracts of fish ovaries produce a progestational effect in mammals and a small amount of estrogenic material The chemical nature of the testicular hormone of fish has not been established but it is reasonably certain that at least some fish produce androgens

Amphibians and reptiles Hypophysectomy interrupts the sexual cycle and spermatogenesis of frogs and induces degenerative changes of the ovary within a few weeks in the *Amblystoma* Female frogs and toads may be induced to ovulate at any time of the year, except immediately after an egg laying period, by injection of macerated anterior pituitary Preparations of pituitaries from females are about twice as effective as those from males Injection of pituitary extract into male frogs induces copulatory reflexes, and the induction of ovulation in *Xenopus* in captivity by the administration of urine from a pregnant woman, which contains gonadotropic substances, has been used as a pregnancy test by L Hogben. It is generally agreed that the development of secondary sexual characters in the amphibians is controlled by sex hormones secreted by the gonads Although it is improbable that the corpora atretica in the amphibian ovary arise by pituitary stimulation, it may be that follicle cells in these bodies respond to the pituitary by secreting sex hormones A simple form of corpus luteum containing yellowish flocculent pigment arises after ovulation, but there is no evidence of any functional importance In some viviparous snakes, however, indications of luteal functions have been adduced Ovariectomy during early gestation, removal of corpora lutea and hypophysectomy all resulted in resorption of embryos or abortion of dead young

In the newt, *Triton cristatus*, there is no interstitial glandular tissue in the testes until the approach of the breeding season This transformation is temporarily associated with the differentiation of the seasonal secondary sexual adornments in the male Destruction of the interstitial glandular tissue is equivalent to castration in preventing development of these secondary characters In the reptile the secondary male sexual characters—skin color, development of dorsal crests, spine, femoral and preanal pores, and a bulge at the root of tail—are associated with the development of the interstitial cells of the testis and also show seasonal variations Castration induces an obvious change of sexual character, but injection of androgen induces hypertrophy of the epididymis and vasa deferentia and stimulates spermatogenetic activity as well

Birds In the male bird, the testes of some species can increase several hundred times in size during the breeding season Hypophysectomy leads to the same effects that occur naturally at the end of breeding season in some wild species Cholesterol is considered to be a precursor of steroid hormones in the bird testes The presence of FSH and LH in the bird pituitary has been amply reported Prolactin is considered an important hormone in birds because it is associated with broodiness as well as with crop milk secretion in pigeons. Large amounts of androgen, estrogen, luteoid, and perhaps progesterone are released from bird ovaries Estrogen induces a hundredfold enlargement of chicken oviducts in 10 days, especially when administered in combination with progesterone The comb growth of young chicks is one of the best known bioassays of androgen

Neuroendocrine function. The associated physiological activities of the nervous and of the endocrine system that influence animal reproduction are briefly dealt with here to illustrate the intrinsic mechanism involved In the lower forms of life especially in those forms without a nervous system, the rhythm of reproduction may be controlled metabolically by the direct action of environmental factors food, temperature, light, humidity, and chemical composition of the environment In the higher forms certain external factors act through the intermediation of the nervous system In the bird the number of eggs in a clutch is generally constant within narrow limits, if the eggs are withdrawn shortly after they are laid, many birds will go on laying, making an attempt to lay the right number In the pigeon, ovulation often is induced by courtship with another pigeon The number of eggs laid has been reported to be increased if a mirror is placed in front of a pigeon cage These instances illustrate the influence of the nervous system on reproduction

Stimuli Experimental study on birds by W Rowan and on ferrets by T Bissonnette has shown that breeding can be induced in midwinter by artificial light Hypophysectomized and blind ferrets do not ordinarily react to light as expected, and it is obvious that the stimulus must be passed through the eye optic nerve, or some receptors in the brain region and thence to the anterior pituitary Moreover the rabbit ferret and ground squirrel normally ovulate in response to the stimulation of copulation This stimulation to switch from the follicular phase to the luteal phase can not be effected in the absence of the pituitary but can be brought about by the injection of pituitary extracts or pituitarylike extracts, pregnant woman's urine (PU) or pregnant mare serum (PMS), so it would seem that this stimulus is normally due to nervous reflexes through the pituitary The stimulus however may be carried by several nervous paths, because local anesthesia of vagina and vulva,

implants itself in the uterine wall Should normal descent be delayed for reasons which are usually unknown the implantation occurs wherever the egg lodges Because conditions are unfavorable in abnormal locations the placenta does not develop properly and the fetus usually dies after a few weeks. At this time hemorrhage into the perito neum and rupture of the tube are common catas trophes. Only rarely does an ectopic pregnancy proceed to term.

Uterus The remarkable cyclic function of men strual on is subject to a wide variety of disorders reflecting both local disturbances in the uterus and distant malfunction of the controlling pituitary gland and ovaries The most important disorders are menorrhagia (excessive bleeding during menstrual periods) and metrorrhagia (bleeding between pe riods) The most common cause of these condi ions s fa lure of ovulation so that bleeding occurs rom a thick hyperplastic endometrium stimulated *a/r by follicular* hormones Other common causes

rently gravid uterus either as a complication of bortion or following intrauterine manipulation dur ng delivery

Benign tumors A common tumor is the polyp a projection of *mucosa made* up of numerous glands Also very common particularly in Negro women a e fibro d tumors, indicated by round masses of overgrown muscle and connective tissue fibers These fibro d *tumors are often numerous* large and associated with pressure symptoms or abnormal bleeding.

the chor onic villi of the placenta as *a series of* grapel ke projections In this instance no fetus is found. The mal gnant variant is known as chorion e thel oma and usually has a dramatically rapid course w th widespread hemorrhagic metastases. Both hydat d mole and chorionephithelioma pro duce the same hormone as the normal placenta This i known as chorionic gonadotropin and is the basis of the usual pregnancy tests performed on such animals as toads, rabbits or rats The hormone also may cause breast enlargement and secretion

Pel c endometriosis What appear to be normal endometr al glands and stroma can develop in ab normal locations, *particularly* the outer wall of the uterus, the ovaries and the tissues between the rect m and the vagina This is known as endo metriosis. It is not certain whether this occurs be cause of spillage of *endometrium* through the tubes during men truation or because of metaplasia of the in olved tissues The hemorrhagic cysts formed undergo cyclic changes during menstruation and may be very painful Symptoms *cease at the meno*- pause. See METAPLASIA.

Cervix uteri *Chronic inflammation* or cervicitis is present in almost all adult women When severe a white discharge leukorrhea develops Epidermoid carcinoma of the cervix is of great interest because *it is common and accessible to early diagnosis* by the cytologic smear technique of G N Papani colaou Early development confined to the epithe lium (cancer in situ) apparently precedes invasive *growth by about 7 years and presents an almost* 100% curable stage When untreated the tumor becomes invasive spreads to the vagina and often blocks the ureters

Vagina *The vagina itself is rarely diseased* Weakening of the pelvic supports often resulting from childbearing produces bulges into the vagina Common forms are cystocele sagging of the blad der wall *rectocele sagging of the rectal wall* and prolapse downward protrusion of the uterus into the vagina protruding externally in severe instances

Vulva Rarely diseased before the menopause the vulva undergoes *atrophic changes in older women* Certain forms of extreme atrophy known as leuko plakia and kraurosis are precancerous ultimately resulting in epidermoid carcinomas

MALE SYSTEM

Testicles During fetal life the testicles migrate from their site of formation near the kidneys to the *inguinal region and thence through the inguinal* ca nals to enter the scrotum at about the time of birth Not infrequently one or both testes are arrested within the inguinal canals or abdomen a condition *known as cryptorchidism If the testis does not de*- scend into the scrotum by puberty either sponta neously or following treatment permanent damage to the sperm forming tubular cells results probably *because the temperature is higher in the body than* in the scrotum The hormone forming interstitial cells can develop normally in a cryptorchid testis

Complete absence of the interstitial cell func *tion* either *because of castration or disease oc*- curring before puberty results in eunuchoidism This is manifest by failure of development of sec ondary sex characteristics such as beard male hair distribution and deep voice by lack *of libido* often by increase in size and weight and by a rela tively feminine personality

Male sterility may be caused by obstruction to flow of seminal fluid or more commonly by pro duction of insufficient or poorly developed sperm cells This in turn may result from organic causes such as inflammation or from such functional dis turbances as hypothyroidism or malnutrition Most often no cause is found

Inflammation of the testis orchitis is most often produced by the mumps virus Salivary gland swell ing may or may not be associated Unfortunately the testicular swelling is often followed by atrophy and sterility of the involved testis An occasional cause of swelling and later atrophy is torsion a twisting of the blood vessels

Malignant tumors of the testis are fairly frequent and carry a high mortality rate The seminoma is

fertilization, and the time of mating in the higher vertebrates

3 The number of embryos developing into self sustaining individuals The probability of normal development depends upon the location where ova are deposited and the protection that the parents give to the zygotes in the lower vertebrates In the higher animals, it depends upon the transportation of zygotes to the prepared uterus at the right time, and the physiological activities of embryo and of mother for proper implantation, for the maintenance of pregnancy, and for proper parturition

Fertility and sterility control The problems of the control of fertility and sterility in human beings and in animals is of great importance for human welfare For pest control and for population control, techniques to induce sterility have been devised For increasing the population of useful animals, techniques to improve fertility are of major importance Among these techniques artificial insemination (the collection, storage, and deposition of sperm into the female) is widely practiced in animal husbandry Artificial pregnancy, the recovery of fertilized ova from one animal and their transference to several other animals is a means of increasing good genetic characters of females This has been proved by M C Chang to be successful in the rabbit, but its application to other animals is still uncertain

In spite of its major importance in the areas of human economic and social welfare intensive study of reproduction lagged far behind other physiological inquiries because of various emotional and prejudicial attitudes Since 1930, however, knowledge of reproductive processes has increased immensely but there are still unsolved problems concerning many aspects of reproduction Application of modern biological chemical, and physical techniques however will advance understanding and control of reproductive processes *See* REPRODUCTIVE BEHAVIOR [M C C]

Bibliography W Andrew *Textbook of Comparative Histology*, 1959, R I Dorfman and R A Shipley *Androgens Biochemistry, Physiology and Clinical Significance*, 1956, W R Lyons et al The hormonal control of mammary growth and lactation, *Recent Progr Hormone Research*, 14 219–248 1958 R C Merrill Estriol A review, *Physiol Revs*, 38 463 1958 C D Turner *General Endocrinology*, 2d ed 1955, C K Weichert, *Anatomy of the Chordates*, 2d ed 1958, B H Willier, P A Weiss and V Hamburger (eds), *Analysis of Development*, 1955

Reproductive system disorders

Those disorders which involve the structures comprising the female and male reproductive organs and accessory structures

FEMALE SYSTEM

Ovaries. If the ovaries are congenitally absent or are lost before puberty the usual feminine characteristics of breast and hip structure, voice, and pubic hair distribution do not develop Destruction after puberty causes cessation of menstruation sterility, and menopausal changes

During active menstrual life, a developing ovum matures each month and is discharged from one ovary. At the time it ruptures from the follicle there is often a small hemorrhage causing a low abdominal pain, known as mittelschmerz, halfway between menstrual periods The developing follicle and the corpus luteum formed after the egg is shed are frequently the site of nonneoplastic cysts, filled with retained secretion or blood, these are not true neoplasms but often cause pain

Neoplastic cysts of the ovary fall into three main categories Dermoid cysts have a thick wall and usually contain hair, teeth, and sebum They apparently represent partial maturation of an ovum and are almost invariably benign Serous cyst adenomas are lined by cells resembling normal Fallopian tube epithelium They contain watery fluid may be moderately large, and are only occasionally malignant Pseudomucinous cystadenomas are lined by tall mucin producing cells and may become filled with large amounts of mucoid fluid Usually they are benign, but if they rupture or if the lining cells develop on the outside wall, there may be widespread implants in the peritoneal cavity

Solid tumors of the ovary present a great variety of benign and malignant forms, some of which are fairly common Many arise from hormone producing cells and exert hormonal effects The granulosa cell tumor and theca cell tumor cause feminization, manifested by abnormal uterine bleeding at any age and by precocious sex development if they occur before puberty The arrhenoblastoma produces male sex hormone inducing virilism, with growth of beard development of deep voice, and cessation of menstruation Carcinomas of the ovary, whether hormone producing or not, and whether solid or partly cystic, have a high mortality Ascites is almost always present, early and widespread metastases are frequent

Oophoritis inflammation of the ovary, is usually a complication of salpingitis *See* GONORRHEA

Fallopian tubes (oviducts). Salpingitis, inflammation of the tubes, is usually bilateral and originates in infection ascending from the uterus The two major causes are gonorrhea and infected abortions the latter usually induced by unclean instruments If the inflammation is severe or repeated the outer ends of the tubes become sealed and the tube enlarges and fills with pus, this is known as pyosalpinx In a later stage the pus disappears and is replaced by a thin watery fluid in the condition called hydrosalpinx Often the adjacent ovary is involved in salpingo oophoritis Sterility commonly results because the tube is permanently closed

Ectopic pregnancy or extrauterine pregnancy is usually found in the Fallopian tube although abdominal and ovarian implantations may occur occasionally Normally the sperm cell ascends the uterus, enters the tube, and fertilizes the ovum high in the tube The fertilized ovum then descends and

mplants itself in the uterine wall Should normal lescent be delayed for reasons which are usually unknown, the implantation occurs wherever the egg lodges. Becau e conditions are unfavorable to abnormal locations the placenta does not develop properly and the fetus usually dies after a few weeks. At this time hemorrhage into the peritoneum and rupture of the tube are common catastrophes. Only rarely does an ectopic pregnancy proceed to term.

Uterus. The remarkable cyclic function of menstruation is subject to a wide variety of disorders

menorrhagia (excessive bleeding during menstrual periods) and metrorrhagia (bleeding between periods) The most common cause of these conditions is failure of ovulation so that bleeding occurs from a thick hyperplastic endometrium stimulated only by follicular hormones Other common causes of abnormal bleeding include abortions and benign and malignant tumors. Endometritis, inflammation of the uterine lining occurs almo t exclusively in a recently gravid uterus, either as a complication of abortion or following intrauterine manipulation during delivery

Benign tumors A common tumor is the polyp, a projection of mucosa made up of numerous glands Also very common particularly in Negro women are fibroid tumors indicated by round masses of overgrown muscle and connective tissue fibers. These fibroid tumors are often numerous large and associated with pressure symptoms or abnormal bleeding.

it invades locally although it may also metastasize A unique tumor is the hydatid mole arising from the chorionic villi of the placenta as a series of grapelike projections In this instance no fetus is found. The malignant variant is known as chorionepithelioma and usually has a dramatically rapid course with widespread hemorrhagic metastases Both hydatid mole and chorionephithelioma produce the same hormone as the normal placenta

also may cause breast enlargement and secretion

Pelvic endometriosis What appear to be normal endometrial glands and stroma can develop in abnormal locations particularly the outer wall of the uterus, the ovaries and the tissues between the rectum and the vagina This is known as endometriosis It is not certain whether this occurs because of spillage of endometrium through the tubes during menstruation or because of metaplasia of the involved tissues The hemorrhagic cysts formed undergo cyclic changes during menstruation and may be very painful Symptoms cease at the menopause. See METAPLASIA.

Cervix uteri. Chronic inflammation or cervicitis is present in almost all adult women When severe a white discharge, leukorrhea, devel ops Epidermoid carcinoma of the cervix is of great interest because it is common and accessible to early diagnosis by the cytologic smear technique of G N Papanicolaou Early development confined to the epithelium (cancer in situ) apparently precedes invasive growth by about 7 years and presents an almost 100% curable stage When untreated the tumor becomes invasive, spreads to the vagina and often blocks the ureters

Vagina. The vagina itself is rarely diseased Weakening of the pelvic supports often resulting from childbearing produces bulges into the vagina Common forms are cystocele sagging of the bladder wall, rectocele, sagging of the rectal wall, and prolapse downward protrusion of the uterus into the

MALE SYSTEM

Testicles During fetal life the testicles migrate from their site of formation near the kidneys to the inguinal region and thence through the inguinal canals to enter the scrotum at about the time of birth Not infrequently one or both testes are arrested within the inguinal canals or abdomen a condition known as cryptorchidism If the testis does not descend into the scrotum by puberty either spontaneously or following treatment permanent damage to the sperm forming tubular cells results probably because the temperature is higher in the body than in the scrotum The hormone-forming interstitial cells can develop normally in a cryptorchid testis

Complete absence of the interstitial cell function either because of castration or disease occurring before puberty results in eunuchoidism This is manifest by failure of development of secondary sex characteristics such as beard male hair distribution and deep voice by lack of libido, often by increase in size and weight and by a relatively feminine personality

Male sterility may be caused by obstruction to flow of seminal fluid or more commonly by production of insufficient or poorly developed sperm cells This in turn may result from organic causes such as inflammation or from such functional disturbances as hypothyroidism or malnutrition Most often no cause is found

the testicular swelling is often followed by atrophy and sterility of the involved testis. An occasional cause of swelling and later atrophy is torsion a twisting of the blood vessels

Malignant tumors of the testis are fairly frequent and carry a high mortality rate The seminoma is

composed of uniform appearing cells, occurs at all ages, and is fatal in about 50% of the subjects despite any treatment The other tumors, known as embryonal carcinomas and teratocarcinomas, are composed of more primitive appearing cells of varying types, occur chiefly in young men, and are almost always fatal because of distant metastasis

The membranes about the testis may become distended with clear fluid, a common condition termed hydrocele

Epididymis. There are two major causes of acute inflammation causing epididymitis In gonorrhea the organism enters the urethra during sexual contact and ascends the urethra and vas deferens to the epididymis In the nonspecific form the organisms are usually introduced by instrumentation of the urethra or surgery of the prostate Tuberculous epididymitis is a chronic inflammation in which the organisms apparently gain access from the blood stream Any form of inflammation is likely to cause permanent blockage of the ducts

Vas deferens and seminal vesicles. These are rarely the site of disease Surgical interruption of both vasa is done frequently either to produce sterility or to prevent ascending infection prior to prostatic surgery The testicles are not thereby injured

Prostate. Relatively minor ailments include prostatitis, usually nonspecific, and small calculi Probably the most frequent and important disorder is a benign overgrowth for which such synonyms as hyperplasia and hypertrophy are widely used The cause is unknown but the usual occurrence after the age of 60 suggests a relation to hormone balance Symptoms are those of partial obstruction of urine flow in the prostatic urethra and include difficulty and frequency of micturition Because of the resulting obstruction and residual urine infection is common, often ascending to the kidneys Carcinoma arising in the prostatic glands is frequent in old men It spreads primarily to the bones, producing a marked reaction which is usually painful This tumor produces great quantities of an enzyme known as acid phosphatase, also found in normal prostatic tissue and secretions An interesting feature of this tumor is its hormone dependence, its growth being stimulated by male sex hormones and retarded by female hormones Because of this even advanced cases are often greatly benefited by removal of the testicles and administration of female hormones

Urethra. The male urethra is frequently the site of inflammation, known as urethritis Some cases are the result of gonorrhea, being acquired through sexual contact Before the era of chemotherapy gonorrheal urethritis, or the treatment employed for it, often resulted in scars known as strictures, which obstructed urine flow Most urethritis is nonspecific and is caused by a variety of bacteria and viruses, and not of venereal origin

Penis. In many parts of the world the foreskin is often removed either for ritual or for medical reasons If left in place it may be the seat of inflammation (balanitis) or marked narrowing (phimosis) which may even interfere with urination The primary lesion of syphilis, the hard buttonlike chancre, is usually found on either the foreskin or glans penis, after it heals a scar is left A warty growth of the penis, called venereal wart, is frequent in some parts of the world and is probably acquired during sexual intercourse Carcinoma arising from the surface epithelium is fairly common, it does not occur in men circumcised during infancy Spread is by local extension and by metastasis to the lymph nodes of the groin Another relatively common disorder of the penis is Peyronie's disease, a form of scarring of the shaft which may interfere with normal erections, its cause is unknown *See* ENDOCRINE SYSTEM, GONAD, ONCOLOGY, REPRODUCTIVE SYSTEM [R N B]

Bibliography M F Campbell, *Urology*, 3 vols., 1954, E Novak, *Gynecologic and Obstetric Pathology*, 4th ed , 1958

Reptilia

A class of vertebrates composed of four living orders the turtles or Chelonia, the tuatara or Rhynchocephalia, the lizards and snakes or Squamata, and the crocodilians or Crocodilia Numerous extinct orders are also known The group first appeared in the Carboniferous and underwent a culminating evolutionary radiation in the Mesozoic, often called the age of reptiles Although the major portion of the class is now extinct, several recent groups particularly the Squamata, are very successful and there are approximately 5000 living species of reptiles as compared to about 4000 living mammals

Characteristics of Reptilia. The reptiles are the most primitive of the completely terrestrial vertebrates and are consequently the first to exhibit amniote features (*see* AMNIOTA) *The earliest terrestrial class* Amphibia is characterized by retaining the type of development found in fishes Most amphibians although terrestrial as adults, return to the water to breed and their virtually naked eggs undergo development in the water In addition amphibians usually have a free-living aquatic larval stage that breathes by means of gills and that metamorphoses rather suddenly into a lunged terrestrial form In the reptiles the cleidoic eggs are covered by a complex series of protective layers including a leathery or calcareous shell A rich supply of food material in the form of yolk is deposited inside the ovum to furnish food for the developing embryo A series of protective extraembryonic membranes the serosa and amnion, appears later in embryogenesis to protect the embryo from water loss and shock as well as the allantois which functions as a storage sac for nitrogenous wastes The serosa and allantois usually fuse to form a respiratory structure Gaseous exchanges take place across the shell and seroallantoic membrane between the outside air and the blood vessels of the allantois All of these adaptations made it possible for the reptile egg to be deposited on land undergo its development there, and hatch into a fully developed form without a

gilled larval stage Most reptilian eggs are buried in the soil or rotting vegetation out of the direct sunlight.

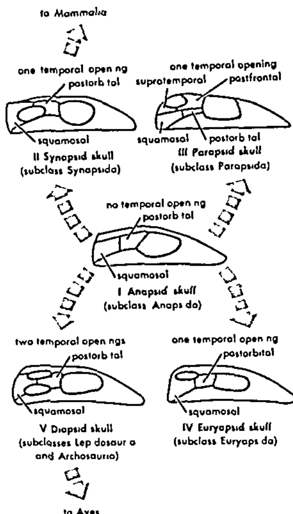
... termediate fossil forms are taken into ac

groups near the evolutionary transition betwee the two classes known only from bony remains, cannot be because in available.

an anamniote egg are so divergent from reptiles that no difficulty is encountered in distinguishing between them (*see* ANAMNIA) All living amphibians have a smooth moist skin, richly supplied with mucous glands two occipital condyles, and many other internal characteristics that separate them from living reptiles, whereas Recent reptiles have the body covered by discrete dry horny scales no mucous glands in the skin, and a single occipital condyle.

Comparison of Reptilia. The Mammalia which are descended from one reptilian stock differ markedly from the Reptilia as indicated below

Reptilia	*Mammalia*
1 No hair	1 Hair, at least embryonically
2 No mammary glands	2 Mammary glands
3 Dependent *upon environmental* heat sources temperature regulated by gross behavioral adjustments, body temperature fluctuating (ectothermic)	3 *Heat produced by body* metabolism, physiological temperature regulatory mechanisms, body temperature rather constant (endothermic)
4 Lower jaw composed of several bones	4 Lower jaw composed of a single element (dentary)
5 Middle *ear region* with a single ossicle (stapes) if functional, transmitting vibrations to inner ear	5 Middle ear with three ossicles (stapes incus, malleus) transmitting *vibrations to inner ear*
6 Heart composed of two auricles and a single partially divided ventricle (fully divided in Crocodilia)	6 Heart composed of two auricles and two ventricles
7 A pair of systemic arteries	7 A single systemic artery
8 Pleural cavities and peritoneal cavities continuous (except in crocodilians where separated)	8 Pleural and peritoneal *cavities separated by a* muscular diaphragm
9 Primarily oviparous some lizards and snakes ovoviviparous or viviparous (sero allantoic placenta)	9 Primarily viviparous (seroallantoic placenta) a few oviparous (monotremes) some ovoviviparous (marsupials)



Lateral views of principal reptilian skull types

*The class Aves is also descended from a rep*tilian stock but no difficulty is encountered in separating the two classes since birds are covered by feathers are endothermic, and have a four chambered heart and a single systemic artery

Phylogeny. All reptiles living and fossil are divided into six major groups based primarily upon the characteristics of the skull These are the subclasses Anapsida Synapsida Parapsida Euryapsida Lepidosauria and Archosauria

Most of the living forms are diapsid lepidosaurians but crocodilians are archosaurians and the turtles belong to the Anapsida Birds are derived *from archosaur ancestors* while mammals evolved from certain synapsid forms *See* PLACENTATION, VERTEBRATA [J M S]

Bibliography A d A Bellairs, *Reptiles* 1957, C H Pope *The Reptile World*, 1955, A S Romer *The Osteology of the Reptiles*, 1956, M A Smith *The British Amphibians and Reptiles*, 1951

Reptilia fossils

Reptilia constitute an order of both living and fossil forms of the phylum Chordata Reptiles were the dominant vertebrate group in the Mesozoic, which is known as the Age of Reptiles Among the better

Representative classification scheme of class Reptilia

Taxonomic group	Common name	Example
Subclass Anapsida	Anapsids	
Order Cotylosauria	Stem reptiles	*Labidosaurus Desmalodon*
Order Chelonia	Turtles	*Chelonides Archelon*
Subclass Ichthyopterygia	Fishlike reptiles	
Order Ichthyosauria	Ichthyosaurs	*Ichthyosaurus*
Subclass Synaptosauria		
Order Protosauria	Protosaurs	*Araeoscelis*
Order Sauropterygia		
Suborder Nothosauria	Nothosaurs	*Nothosaurus*
Suborder Plesiosauria	Plesiosaurs	*Plesiosaurus*
Suborder Placodontia	Placodonts	*Placodus*
Subclass Lepidosauria		
Order Eosuchia	Eosuchians	*Youngina*
Order Rhynchocephalia	Rhynchocephalians	*Homeosaurus*
Order Squamata		
Suborder Lacertilia (Sauria)	Lizards mosasaurs	*Tylosaurus*
	Snakes	*Palaeophis*
	Thecodonts	*Ornithosuchus*
Order Crocodilia	Crocodilians	*Protosuchus Stenosaurus*
Order Pterosauria	Flying reptiles pterosaurs	*Pterodactylus Pteranodon*
Order Saurischia	Reptilelike dinosaurs	
Suborder Theropoda	Theropods	*Tyrannosaurus Palaeosaurus*
Suborder Sauropoda	Sauropods	*Brontosaurus Diplodocus*
Order Ornithischia	Birdlike dinosaurs	
Suborder Ornithopoda	Duck billed dinosaurs	*Camptosaurus Hypsilophodon*
Suborder Stegosauria	Stegosaurs	*Stegosaurus*
Suborder Ankylosauria	Ankylosaurs armored dinosaurs	*Nodosaurus Ankylosaurus*
Suborder Ceratopsia	Horned dinosaurs	*Triceratops Protoceratops*
Subclass Synapsida	Synapsids	
Order Pelycosauria	Pelycosaurs	*Dimetrodon*
Order Therapsida	Therapsids	*Cynognathus Dicynodon*
Order Ictidosauria	Ictidosaurians	*Microconodon*

known fossil reptilians are the cotylosaurs dinosaurs placodonts thecodonts synapsids and pelsiosaurs Reptiles exhibited a great diversity in form and they were to be found in all types of habitats Adaptation to a terrestrial existence was a significant evolutionary development among the reptiles This was realized through the production of eggs protected by an outer covering and capable of development on land as well as by a modification of the dermis which prevented dehydration The many prevalent species of ancient reptiles disappeared suddenly and it has been postulated that climatic changes which affected the habitat or the organisms themselves or both caused the decline in the number of species

Fossil groups There are divergences of opinion on the systematic status of the various groups which comprise the Reptilia The accompanying table represents one scheme of classification and lists the common name for the group and a fossil example

Paleoecology Fossil reptiles inhabited swamps fresh water and marine habitats and land Some species were predaceous carnivores others were either herbivores or omnivores According to F Broili the terrestrial *Homoeosaurus* fed on marine fishes and was occasionally devoured in turn by larger fish Morphological studies of mosasaurs indicate that some species were diving forms and others were surface swimmers They fed on fish mollusks and echinoderms Terrestrial species fed on smaller reptiles or vegetation or both *See* PALEOBOTANY PALEOCLIMATOLOGY PALEONTOLOGY, REPTILIA [C B.C]

Repulsion motor

An alternating current (ac) commutator motor designed for single phase operation The chief distinction between the repulsion motor and the single-phase series motors is in the way the armature receives its power In the series motor the power is supplied by conduction from the supply circuit and in the repulsion motor by induction from the field of the stator winding For discussion of the ac series motor *see* UNIVERSAL MOTOR *see also* ALTERNATING CURRENT MOTOR

The repulsion motor has a primary or stationary field winding which is connected to the power supply and a secondary or armature winding which is mounted on the motor shaft and rotates with it The terminals of the armature winding are short circuited through a commutator and brushes There is no electrical contact between the field and armature (Fig 1) *See* WINDINGS (ELECTRIC MACHINERY)

If the motor is at re t and the field coils are energized from an outside ac source a current is in

duced in the armature just as in a static transformer If the bru hes are in line with the neutral axis of the magnetic field there is no torque or tend

sis ance in series with the motor as is done with direct-current series motors. A more common method is to start the motor with reduced voltage and increase the voltage as the motor increases speed. This can be conveniently done with a transformer with an adjustable tapped secondary

It is also possible to doubly feed the motor that is, the armature may receive its power not only by induction from the stator winding but also by conduction from a transformer with adjustable taps as in Fig. 2.

Repulsion start, induction run motor This motor possesses the characteristics of the repulsion motor at low speeds and those of the induction motor at high speed (*see* INDUCTION MOTOR) It starts as a repulsion motor At a predetermined speed (generally at about two-thirds of synchronous speed) a centrifugal device lifts the brushes from the commutator and short circuits the armature coils The motor then runs as an induction motor In Fig 3*b* the curve *AB* represents the characteristics of an induction motor and curve *CD* a repulsion motor The solid curve *AD* is the combined characteristic of a repulsion start induction run motor

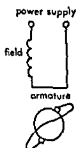


Fig 1 Schematic of a repulsion motor

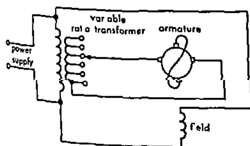


Fig 2 Armature excited repulsion motor

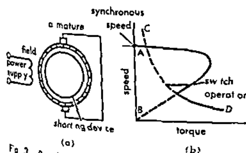


Fig 3 Repulsion-start induction run motor (a) Schematic diagram (b) Speed torque characteristic

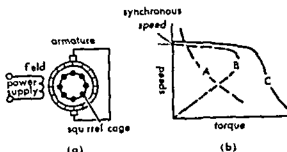


Fig 4 Repulsion induction motor (a) Schematic diagram (b) Speed torque characteristic

Repulsion induction motor. This motor is very similar to the standard repulsion motor in construction except for the addition of a high resistance squirrel cage on the rotor Both rotor windings are torque-producing and the total torque produced is the sum of the individual torques developed in these two windings In Fig 4*b* curve *A* is the characteristic of the repulsion motor torque developed in this motor Curve *B* represents the induction motor torque Curve *C* is the combined total torque of the motor The advantages of this machine are its high starting torque and good speed regulation Its disadvantages are its poor commutation and high initial cost [S W]

Bibliography A F Puchstein T C. Lloyd and A C Conrad *Alternating Current Machines*, 3d ed, 1954

Reserve battery

A battery which is inert until an operation is performed which brings all the cell components into the proper state and location to become active Several types have been developed In water activated or electrolyte activated batteries the water or electrolyte component is not present during storage It is added just before the cell is put into use In thermal batteries the electrolyte is a solid at room temperature and has very low conductivity By raising the temperature above the melting point the conductivity of the electrolyte becomes excellent and the cell is capable of delivering significant power

Water activated batteries Practical battery systems have been developed using magnesium anodes against silver chloride or cuprous chloride cathodes Cuprous chloride cathodes are less expensive than silver chloride cathodes but they are also bulkier and less stable particularly in a humid atmosphere

The batteries are assembled dry The active elements may be separated by porous paper or other

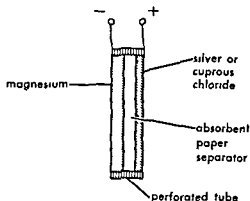
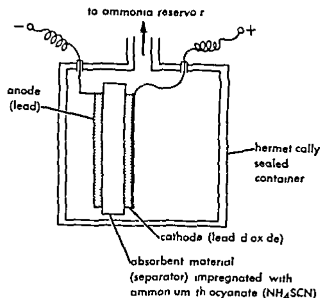


Fig 1 Schemat c of water-activated cell



F g 2 Schemat c of ammonia vapor-activated reserve type primary cell

inert media Water may be poured into a container holding the elements or may flow continuously through the element Either fresh or salt water may be used

Cells with absorbent separators can be activated by immersion Subsequent operation may either be in air using only the water retained in the cells or while immersed Performance of a 2 cell battery immersed in sea water showed an output of 18 watt hours/pound (whr/lb) when fully discharged in 6 min

The dry elements stored in a sealed container are capable of indefinite storage life

Electrolyte activated batteries Any cell can be made as a reserve electrolyte cell If the electrodes are in place it is necessary only to add the electrolyte to make a complete cell In practice however the separation of the electrolyte is done only when excessive deterioration would occur during wet storage prior to use Great ingenuity has been shown in designing complete battery packages in which an aqueous electrolyte is stored in a separate chamber The package contains a mechanism which may be operated from a remote location which drives the electrolyte out of the reservoir and into all the cells of the battery In general these packaged batteries have been used only in military applications.

Gas activated batteries It is also practicable to activate a battery with a gas The gas combines with a dry salt in the battery to form a liquid having good conductivity Boron trifluoride gas will react with dry hydrated barium hydroxide to form a highly acid solution containing barium salts, borates and fluoborates Ammonia gas reacts with ammonium salts to form a solution having good conductivity These gas activated batteries are reported to operate well over a wide temperature range

Thermal batteries Some compounds such as sodium chloride and potassium hydroxide, show very low conductivity in the solid state at room temperature but very good conductivity in the molten state For example a mixture of sodium hydroxide and potassium hydroxide becomes an excellent ionic conductor when heated above 170°C. By combining a zinc anode and a silver oxide cathode with solid pads of the eutectic mixture, all the elements of a cell are present The electrolyte has an appreciable amount of entrained moisture, which plays a role in the discharge but the cell will also work with carefully dried materials Such cells are capable of high power output for a few minutes when heated to 200°C or higher At 1 amp/in ^{2} of positive plate the cell voltage is 1 16 at 200°C, 1 23 at 250°C, 1 30 at 300°C.

Thermal batteries are capable of operation at very low ambient temperatures provided that a suitable heat source is available to melt the electrolyte For ordinary temperatures they are not advantageous as compared with reserve aqueous electrolyte types

A magnesium and manganese dioxide cell with sodium hydroxide electrolyte can operate for longer discharge times than the zinc–silver oxide cell mentioned previously because of the greater stability of the reactants at high temperatures

Thermal batteries are essentially of low energy efficiency because the heat absorbed in melting the electrolyte is not available in the electrical output. Consequently their use is restricted to small cell sizes [S.EI]

Bibliography W J Hamer and J P Schrodt Investigations of galvanic cells with solid and molten electrolytes *J Am Chem Soc.*, 71 2347 1949 J P Mullen and P L. Howard Character

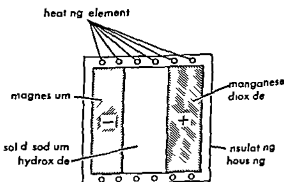


Fig 3 Schemat c of thermal cell

istics of the silver chloride-magnesium water acti ... Trans Electrochem Soc, 90 529,

405 1946, G W Vinal, Primary Batteries, ...
U.S. Army Signal Laboratories, Proceedings Elev- ... of Battery Research and Development

90 467, 1946

Reservoir

A pond or lake built for the storage of water, usually by the construction of a dam across a river See DAM, WATER SUPPLY ENGINEERING. The size of reservoir needed is a function of the water demands, the natural flows of the river impounded, and the extent of droughts to be encountered In areas of moderate rainfall, storage may be required during only a few days or weeks in a year In arid and semiarid areas, storage capacity may be needed to supplement low stream flows over periods of many months or even years In some instances water can be diverted by gravity flow or by pumping from one river to a reservoir on another stream Water supplies may be taken directly from the reservoirs, or water may be released from the reservoirs to augment river flows past a water supply intake downstream. [R H]

Resilience

The ability of a mechanical part to bend under stress and to return to its original position when the force ceases The resilience of the part is the energy that it can deliver in returning to its original position after an elastic deflection A spring is resilient. For comparison of materials a modulus of resilience is defined as the maximum energy in inch pounds stored in a cubic inch of a material when stressed to its elastic limit This modulus equals the area under the stress-strain curve up to the elastic limit See SPRING (MECHANICAL), STRESS AND STRAIN [W J KR]

Resin

An organic polymer Resin, together with filler plasticizer, and other materials, makes up a plastic. Resins are formed either by addition (vinyl type) in which an olefinic compound opens up a double bond to allow chain formation with other ...

... polymers such as polyethylene and polyacrylates are thermoplastic and can be shaped by heating Condensation polymers derived from phenols and aldehydes (phenolic), from polyhydric alcohols and dicarboxylic acids (alkyds, polyesters), or from amino compounds and aldehydes tend to be thermosetting when heated See NAVAL STORES, PLASTICS FABRICATION, POLYMER [A L H]

Resistance, electrical

A measure of the difficulty with which electric current flows through a medium It is defined as the ratio of the potential drop V across a circuit element to the current I flowing in the element. Thus $R = V/I$ If V is in volts and I is in amperes, R is in ohms The resistance of a given piece of wire depends not only upon the material from which it is made, but also upon its length and diameter and upon the frequency of the current At high frequencies the current flows in a thin layer near the surface of the wire Thus, for a given wire, the ac resistance is higher than the dc resistance, and increases with increasing frequency See EDDY CURRENT, SKIN EFFECT

A more fundamental concept, which is independent of geometry, is that of resistivity This depends only upon the material, the ambient temperature, and the pressure Resistivity is defined as the resistance measured between opposite faces of a unit cube of material See RESISTIVITY, ELECTRICAL; see also CURRENT, ELECTRIC, OHM'S LAW [J W ST]

Resistance heating

The generation of heat by electric conductors carrying current The degree of heating for a given current is proportional to the electrical resistance of the conductor If the resistance is high, a large amount of heat is generated, and the material is used as a resistor, rather than as a conductor See RESISTANCE, ELECTRICAL.

Direct heating. When heating metal strip or wire continuously the supporting rolls can be used as electrodes and the strip or wire can be used as the resistor

In Fig 1, the electric current passes from the roll A through the strip or wire to the roll B Heating by this method can be very rapid Disadvantages are that the electric currents are large, and uniform contact between the strip and the rolls is difficult to maintain since both surfaces must be clean and free of oxides For these reasons, direct heating is not used extensively

Ovens and furnaces. If the resistor is located in a thermally insulated chamber, most of the heat generated is conserved and can be applied to a wide variety of heating processes Such insulated cham-

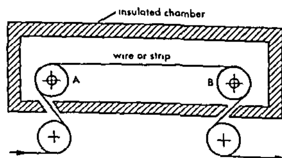


Fig 1 Schematic illustration of direct heating of a metal strip or wire

bers are called ovens or furnaces, depending on the

wool or glass wool between the inner and outer steel casings for thermal insulation Typical uses are for baking or roasting of foods, drying of paints and organic enamels, baking of foundry cores, and low temperature treatments of metals

The term furnace generally applies to units for operation above 1000°F In these, the thermal insulation is made up of an inner wall of fireclay, Kaolin, or high alumina or zirconia brick, depending on the temperature, with secondary insulating blocks made from such base materials as rock wool asbestos fiber, and diatomaceous earth Typical uses of furnaces are for heat treatment of metals melting of metals, vitrification and glazing of ceramic wares, annealing of glass and roasting and calcining of ores See HEATING, ELECTRIC

Actually ovens and furnaces overlap in temperature range, with ovens being used at temperatures as high as 1000°F, and furnaces as low as 250°F Electrically heated ovens and furnaces have advantages over fuel fired units These advantages often compensate for the generally higher cost of electric energy The main advantages are (1) ease of distributing resistors, or heating elements, to obtain a uniform temperature in the product being heated (2) ease of operation, since adjustments by operators are usually unnecessary, (3) cleanliness, (4) comfort, since heat losses are low and there are no waste fuel products, (5) adaptability to the use of controlled furnace atmospheres or vacuum, and (6) high temperatures beyond the range attainable with commercial fuels

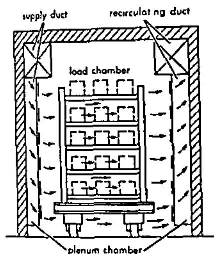
Resistor materials In addition to having high resistivity, heating elements must be able to withstand high temperatures without deterioration or sagging Other desirable characteristics are low resistance change from low to high temperatures formability, low cost, and availability of materials Materials which meet some or all of these requirements are listed in the table

Molten salts The electrical resistance of molten salts between immersed electrodes can be used to generate heat Limiting temperatures are dependent on decomposition or evaporization temperatures of the salt Parts to be heated are immersed in the salt Heating is rapid and, since there is no exposure to air, oxidation is largely prevented Disadvantages are the personnel hazards and discomfort of working close to molten salts

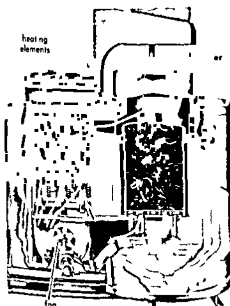
Heat transfer in ovens and furnaces Heating elements mounted in ovens and furnaces may be located to radiate directly to the parts being heated or may be located behind baffles or walls so that direct radiation cannot take place. The heat then is transferred by circulating the furnace air or gas Determination of which method or combination of methods to use is based on temperature uniformity, speed of heating, and high temperature strength limitations of fans or blowers At temperatures below 1200°F radiation is slow, and virtually all ovens and furnaces in this temperature range use forced convection or circulation From 1200°F to 1500°F radiation is increasingly effective while reduced gas density and lower fan or blower speeds (because of reduced strength at high temperatures) makes forced convection less effective Therefore, in this temperature range, combinations of radiation and forced convection are used Above 1500°F forced convection is employed only when direct radiation cannot reach all parts of the loads being heated An example of this is a container filled with bolts or a rack filled with gears

Electric furnace resistor materials and temperature ranges

Material	Maximum temperature °F		Characteristics
	In air	In nonoxidizing atmospheres	
34% Nickel 18% chromium 48% iron	1800°	2100°	Forms protective coating of chrome oxide
60% Nickel 15% chromium 25% iron	1800°	2100°	Forms protective coating of chrome oxide
80% Nickel 20% chromium	2100°	2150°	Forms protective coating of chrome oxide
25% Chromium 6% aluminum 69% iron	2350°	Not used	Forms protective coating of aluminum and chrome oxides
50% Nickel 50% iron	Not used	2250°	Usable only in nonoxidizing atmospheres
Silicon carbide	2800°	2500°	Oxidizes in service to silicon dioxide
Platinum	2900°	2900°	
Molybdenum	Not used	Approx 3400°	gon or in vacuum at max temp
Tungsten	Not used	Approx 3700°	Usable only in hydrogen helium argon or in vacuum at max temp
Carbon	Not used	Approx 5000°	Usable only in helium argon or in vacuum



Fg. 2 Typ cal oven for maximum temperature of 800 F Heat ng is by forced convect on (Carl Mayer Co p)



Fg 3 Low temperature p t type furnace for temper ng steel us ng forced convection (Lindberg Industrial Co p.)

Obviously in vacuum furnaces heat transfer can be only by radiation

Ovens Figure 2 shows a typical oven, for a maximum temperature of 800°F the heating is by forced convection The heated air enters through the supply duct passing downward into the plenum chamber and distributing louvers into and across the load chamber returning through the right side plenum to the recirculating duct The external heater and blower which complete the circuit are not shown in the illustration

Low temperature furnace Figure 3 shows a pit type furnace used for tempering steel It operates at temperatures of 250-1400°F and uses forced convection

The steel parts to be tempered are placed in the load chamber at the right The electrical heating elements are in the heating chamber at the upper left The fan, at the lower left, circulates air upward over the heating elements and into the load chamber The hot air is forced down through the load and back to the fan The heating chamber is thermally insulated from the load chamber so there is no direct radiation Such furnaces are designed commercially to hold the load temperature within 10° of the control temperature, and can be designed for closer control if desired

High temperature furnace Figure 4 shows the interior of a pit type carburizing furnace, which uses radiant heating elements in the form of corrugated metal bands mounted on the inside of the brick walls The work basket (not shown) rests on the load support at the bottom These elements are designed to operate at low voltage (approximately 30 volts) so that soot deposits from the carburizing gases will not cause short circuits *See* FURNACE CONSTRUCTION

Electrical input Electric ovens and furnaces are rated in kilowatts (kw) The electrical input is determined from the energy absorption Q of the load and the thermal losses

The average rate of energy flow into load during the heating period is Q/t, where Q is in kilowatt

temperature approaches furnace temperature Therefore the energy input must be high enough to

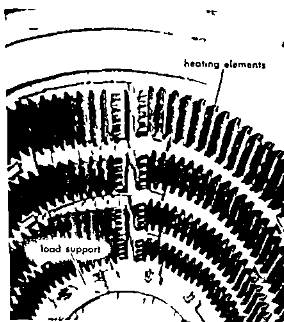


Fig 4 Inter or of high temperature pit type carburiz ing furnace using radiant heating (Lindberg Industrial Corp)

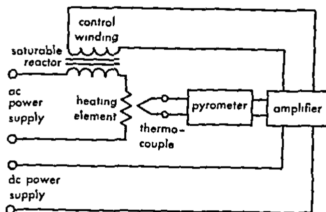


Fig 5 Temperature control using saturable reactor

take care of the high initial heating rate This leads to the following approximate formulas for the input, in which L is the thermal losses in kilowatts (kw) at the operating temperature

For batch furnaces, input in kw = $L + 1\,5\ Q/t$
In continuous furnaces, input in kw =
$L + 1\,25\ Q/t$

Operating voltage for heating elements Heating elements are usually designed to operate at standard service voltages of 115, 230, or 460 volts, if two conditions can be satisfied These are, first, that the heating element is sufficiently heavy in cross section to avoid sagging or deformation in service, and second, that there is no appreciable electrical leakage through the furnace refractories tending to short circuit the heating elements The latter consideration generally limits voltages to 260 volts at 2100°F, and to approximately 50 volts at 3100°F, because the refractory walls of the furnace become increasingly better conductors at higher temperatures

In vacuum furnaces, voltages are limited to 230 volts by the tendency to break down into glow discharge at low pressure

Temperature control. Almost all commercial electric ovens and furnaces have automatic temperature control The simplest control uses magnetic contactors, which open and close the circuit to the heating elements in response to temperature signals from control thermostats or thermoelectric pyrometers A refinement of this ON and OFF control is to modulate it with a timer, with the ON period becoming a progressively smaller percentage as the control temperature is approached This prevents the overshooting which results from thermal lag in the control thermocouple and the furnace

True proportioning control is achieved through the use of saturable reactors in series with the heating elements Figure 5 shows schematically the arrangement used

The pyrometer, through the amplifier controls the direct current to the control winding of the reactor When the direct current is maximum, the reactor offers virtually no impedance to the alternating current flowing through the heating element and the normal amount of heat is generated. As control temperature is approached, the direct current is decreased, increasing the impedance of the reactor and reducing the current to the heating elements until equilibrium is reached *See* PYROMETER, SATURABLE REACTOR, TEMPERATURE CONTROL, AUTOMATIC [W.R.]

Bibliography· W H McAdams, *Heat Transmission*, 3d ed, 1954, M H Mawhinney, *Practical Industrial Furnace Design*, 1928, W Trinks, *Industrial Furnaces*, vol 1, 4th ed, 1951, and vol 2, 3d ed, 1955

Resistance measurement

The quantitative determination of that property of an electrical conductor called electrical resistance. The practical unit of measurement is called the ohm Various engineering applications require that resistance be measured over the range of 0 1 microhm (10^{-7} ohm) to 10,000 megamegohms (10^{16} ohms) *See* OHM'S LAW, RESISTANCE, ELECTRICAL, RESISTIVITY, ELECTRICAL

Resistance may be measured using either direct or alternating current (dc or ac) If dc is used, the true resistance of the conductor is measured When the measurement is made with ac, the result is usually called the effective resistance Physical factors which must be considered when measuring resistance include (1) the ambient temperature and self heating of the conductor, (2) the connecting lead and contact resistance when measuring low resistance conductors, (3) parallel leakage paths around a high resistance conductor, (4) current distribution within the conductor, (5) thermally, electrolytically, and other spuriously generated voltages within the resistance being measured or in the measuring circuit, and (6) the capacitance and inductance associated with the conductor

Practically all methods for measuring resistance are based on Ohm's law, and countless variations of electrical networks have been devised for specific resistance measurement requirements Ohm's law, first proven experimentally in 1826, states that, for dc circuits, the difference of potential E existing between two terminals of a conductor is directly proportional to the current I flowing in the conductor, or $E = RI$ The constant of proportionality R is called the resistance of the conductor (see Fig 1) This relationship is valid for most dc circuits although certain exceptions are found as in gaseous conduction, thermionic conduction, and semiconductors which exhibit nonohmic characteristics Conductors of this type must be measured under completely specified physical conditions per

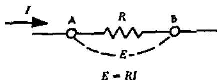


Fig 1 Ohm's law

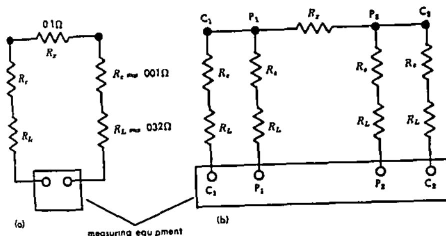


Fig 2 (a) Contact res stance R_c and lead resistance R_L included in measurement of unknown resistance (b) R and R_L effect vely eliminated

tinent to the phenomena, such as gas pressure emission temperature, applied voltage and polarity For a discussion of resistance in a dc circuit see DIRECT CURRENT CIRCUIT THEORY

Ohm s law is also valid for ac circuits if vector quantities are used for all parameters, that is E = ZI where Z is the complex impedance $R_e + jX_e$ R_e is the effective resistance, and X_e is the effective reactance See ALTERNATING-CURRENT CIRCUIT THEORY

Measurement of low resistance. The measurement of low resistance values (10^{-1}–10 ohms) or the accurate measurement of intermediate resist ance values (1 10^4 ohms) requires special tech

... would result Since the test leads and their contact connection resistances are in series with the resistor to be measured their resistance is also included in the measurement

The technique usually employed to eliminate this effect is to provide the unknown resistance with four leads as in Fig 2b Two of these are current connections which supply current to the resist ance the other two are potential connections physically located between the current connections Spe cial measuring circuits requiring little or no cur rent in the potential circuit are used These tech niques allow measurement of only that resistance between the potential contacts See KELVIN BRIDGE

Measurement of high resistance The measurement of high resistances (10^5 to 10^{14} ohms) requires consideration of spurious leakage paths which may exist in parallel with the resistance terminals as shown in Fig 3a Electrical leakage is usually the result of inadequate moisture sensi t e dirty or deteriorated insulation between the terminals. Errors of several thousand per cent may result unless suitable precautions are taken Under extreme conditions such erratic readings may be

...

resistance to be measured Physically a guard consists of a low resistance conductor, electrically insulated from the guarded terminal and located to intercept the leakage current as shown in Fig 3b Measuring circuits designed to take advantage of this technique maintain the guard and the guarded terminals at equal or nearly equal potential to prevent or minimize current flow between the guard and guarded electrodes The guard circuit of the measuring device is also designed to conduct the leakage current around the main measuring circuit so that it has little or no effect on the measurement See INSULATION RESISTANCE TESTING

Measurement of an inductive resistance. For dc measurements of inductive resistance, the time constant of the circuit should be considered (see TIME CONSTANT) The time constant t in seconds is equal to the inductance in henries divided by the resistance in ohms Inductive resistance is found in relay coils transformers chokes and similar components

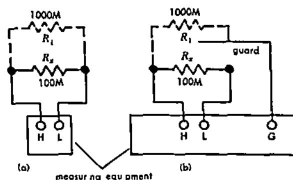


Fig 3 (a) Leakage res stance R_l included in measurement (b) R_l eliminated by guard

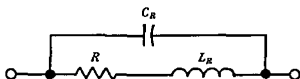


Fig 4 Simple equivalent circuit of a conductor

Upon closing the test potential circuit, the current through this type of resistance requires approximately $6t$ to reach 99% of its steady state value During this time transient conditions exist, and the sensitive detectors of some circuits may be damaged if the detector is connected into the circuit prior to $5t$ or $6t$ Upon opening the supply circuit, the high voltage generated by the collapsing magnetic field of the inductance can cause serious damage to components by voltage breakdown or high surge currents

AC resistance. The ac or effective resistance of a conductor differs from the dc value by an amount that is a complex function of (1) the self inductance and capacitance of the conductor, (2) the effective mutual inductance and capacitance of nearby conductors and shields, and (3) skin effect These factors are dependent upon the test frequency and waveform It is therefore essential that ac resistance measurements be made under carefully and completely specified mounting, connection, and operating conditions *See* SKIN EFFECT

The simplest equivalent circuit of a conductor is shown in Fig 4 where R is the true resistance, L_R is the true inductance and C_R is the self capacitance of the conductor The impedance Z of this circuit at a radian frequency ω is

$$Z = \frac{R + j\omega[L(1 - \omega^2CL) - CR^2]}{(1 - \omega^2CL)^2 + \omega^2C^2R^2}$$

If both L and C are small, as in a resistor, the approximate effective resistance R_e (neglecting skin effect) is

$$R_e \approx R[1 + \omega^2C(2L - CR^2)]$$

If L is large and C is small, as in an inductance coil,

$$R_e \approx R[1 + 2\omega^2CL]$$

If C is large and L is small, as in a capacitor,

$$R_e \approx R[1 - \omega^2C^2R^2]$$

Complication of this simple analysis with the actual physical conditions of mutual capacitance, conductance, and inductance to the earth or to a shield requires a complex solution

Resistance measurement methods. The methods of measuring resistance may be classified as either deflection methods or comparison methods As implied by the names, deflection methods utilize the deflection of ammeters or voltmeters which may be calibrated in terms of resistance under specific operating conditions, while comparison methods are based upon the use of a calibrated resistor which can be compared to an unknown resistor The fundamental deflection method is known as the voltmeter ammeter method Basic comparison methods are by potential drop, the Wheatstone bridge, and the Kelvin bridge

Voltmeter ammeter method This method of measuring resistance is illustrated in Fig 5 Simultaneous readings of the voltmeter and ammeter are taken, and the unknown resistance R_x is calculated from Ohm's law.

$$R_x = V/A$$

For the voltmeter connection shown in Fig 5*a*, the true resistance R_x is slightly larger than the calculated value, because the ammeter measures the sum of the currents in R_x and the voltmeter For the connection of Fig 5*b*, R_x will be slightly smaller than that calculated, since the voltmeter reading is larger For the most accurate measurements, a correction must be applied as required by the connection Circuit 5*a* causes the least error when R_x is low, circuit 5*b* causes the least error when R_x is high

While the voltmeter ammeter method serves to illustrate a basic principle, the need for simultaneously reading two meters and then calculating the resistance makes the method inconvenient to use except where resistance must be measured while a circuit is maintained in operation A practical and widely used simplification based on this method is used in the ohmmeter *See* OHMMETER

Comparison by potential drop This method, accomplished with the circuit of Fig 6 is a logical development from the voltmeter ammeter method Only one meter is used for the measurement The unknown resistance R_x is connected in series with a standard resistance R_s which may be either fixed

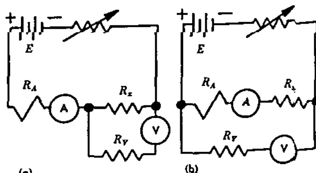


Fig 5 Voltmeter ammeter method for measuring resistance (*a*) Least error when R_x is low (*b*) Least error when R_x is high

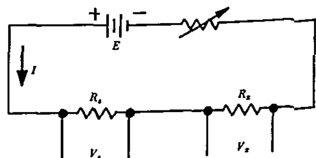


Fig 6 Resistance comparison by potential drop

or adjustable. Since the same current flows in both resistors, by Ohm's law

$$I = V_s/R_s = V_x/R_x$$

or

$$R_x = R_s V_x/V_s$$

If the standard resistance is fixed, R_x may be determined from the ratio of the potential difference readings multiplied by the value of R_s. If the standard is adjustable it may be set to that value for which V_x/V_s equals unity, R_x then equals R_s and may be read directly from the R_s setting. The potential difference may be read with either a voltmeter or a potentiometer. Several readings should be made to insure that the current did not change between successive voltage measurements. This method is seldom used commercially, but it is capable of high accuracy for comparing like-value resistors using a potentiometer for the potential difference measurement. For other comparison methods, see KELVIN BRIDGE; WHEATSTONE BRIDGE.

Resistance standards. High quality standard resistors are used in comparison methods of resistance measurement. The unknown resistance is

that the standard will have a minimum frequency coefficient. See ELECTRICAL STANDARD.

Because of their high degree of permanence resistance standards are used as the basis for comparison to determine the limits of error the preci

than required for the usual engineering determination, resistance standards are often kept and used in the standardizing laboratory to calibrate working instruments, such as bridges meters and less accurate working standards.

Primary resistance standards. These are usually supplied as four-terminal wire-wound or folded strip, manganin resistors in decade values of resistance covering the range of about 10^{-4} to 10^{4} ohms. They are designed for maximum stability as their most important quality, and little attention is given to the ac characteristics of these units. They are usually adjusted to within ±0.01% of their nominal value and a certificate furnished with each unit shows its deviation from nominal in parts per million at the time of measurement under specified conditions of measurement, such as ambient temperature and power dissipation. After several annual or semiannual certifications have been obtained from either the manufacturer or the National Bureau of Standards, an invaluable record of the average drift and random instability will have been accumulated to serve as a guide for estimating the dependable accuracy of the standard with a high degree of certainty.

The highest quality resistance standards ever produced are hermetically sealed wire-wound, 1 ohm resistors designed by J. L. Thomas of the National Bureau of Standards (see Fig. 7a). A group of these standards is used to maintain the reference ohm at the Bureau of Standards, and some of these units have changed less than 1 part per million in several years.

Secondary resistance standards. By virtue of more difficult construction problems or less accurate comparison methods secondary standards can not be guaranteed to as high a degree of accuracy and stability as primary standards. They may be broadly classified as those having values below 10^{-4} ohm and those above 10^{4} ohms, although the intermediate values may also be constructed with less care than required for primary standards and so be classed as secondary standards.

The low resistance secondary standards are always of the four terminal type and are usually made with multiple straight strips of manganin brazed into bus-bar type current connectors (Fig. 8). This construction is necessary to obtain low resistance values and at the same time provide sufficient cooling surface to dissipate the internally generated heat.

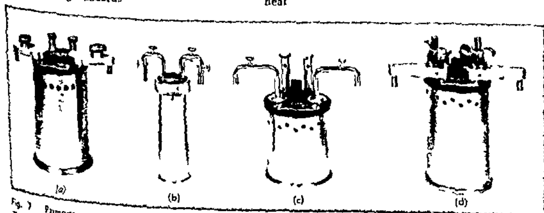


Fig. 7. Primary resistance standards. (a) 1.0-ohm Thomas type. (b) 1.0-ohm Rosa type. (c) 0.1 ohm Reichsanstalt type. (d) 0.001 ohm Reichsanstalt type. (Leeds and Northrup Co.)

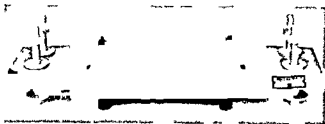


Fig. 8 Secondary resistance standard 0.0004 ohm (*Leeds and Northrup Co.*)

Secondary standards for the higher resistance range (10^4 to 10^7 ohms) are of the wire-wound or woven type. Because of the fine wire required to construct these resistors, they should be hermetically sealed for maximum long-term stability. In addition, the winding supports and the external cases should be treated with a nonwettable material to reduce surface leakage effects under conditions of high humidity.

For *secondary resistance standards above* 10^7 ohms, the carbon film, boro-carbon film, or metallic film units are the best available despite their shortcomings of high temperature coefficients, voltage sensitivity, and relative instability. These units must also be hermetically sealed in glass or ceramic tubes whose surfaces have been treated. Great care must be used in handling and storing these units since the surface of the case must not become contaminated with fingerprints, condensed oil, or chemical vapors if they are to retain their maximum stability.

Resistance decade. This assembly of resistors has a suitable circuit switching device to insert or remove one or more resistors of the circuit and change the total circuit resistance in unit or decade amounts. Although for many applications a resistance decade is used as an adjustable resistance standard, it is fundamentally a standard of resistance change ΔR.

Figure 9 illustrates various switching methods and resistor configurations which have been developed. Regardless of configuration, all resistance decades have a residual and finite "zero" resistance that includes the internal wiring and switch contact resistance. They also have an instability due to *the variation in contact resistance.* For these reasons, the limits of error are usually stated in two factors. The first is a function of the error in the resistors themselves; the second is a function of the residual factors. For low resistance values the residual term may determine the error, while for

ΔR/step	R_s
0.1 ohm	$\sqrt{50.4}$ ohm
0.01 ohm	$\sqrt{5.04}$ ohm
0.001 ohm	$\sqrt{0.504}$ ohm
0.0001 ohm	$\sqrt{0.0504}$ ohm

Fig. 9 (a–e) Resistance decade circuits.

high resistance values it usually can be neglected

Applications There are two major classifications of the applications of resistance measurement those in which resistance is measured because the resistance value must be known and those in which the measured resistance is associated with another physical quantity of interest. The first group would include such measurements as component checking during resistor manufacturing, evaluation of material properties (such as resistivity) for design data, or the testing of equipment for analytical purposes (such as generator winding resistance) Examples falling in the second area are temperature measurement with resistance thermometers insulation resistance for equipment condition check, fault location in cables, strain gages circuit continuity checks, and many others *See* CIRCUIT TESTING, ELECTRICAL, RESISTOR, STRAIN GAGE, THERMOMETER. [C E.A]

Bibliography B Hague *Alternating Current Bridge Methods*, 4th ed, 1938, F K. Harris, *Electrical Measurements*, 1952, F A Laws, *Electrical Measurements* 1938, J L. Thomas, *Natl Bur Standards J Research* 5 295–304, 1930, 36 107–110, 1946 L F Woodruff *Principles of Electric Power Transmission*, 2d ed, 1938

Resistance welding

A general term for a group of methods for joining metals together electrically A low voltage forces a high density current for a relatively short time through an area covered by the welding electrodes A mechanical force must be applied to the electrodes before, *during* and after the time the current flows in order to produce the proper conditions for heating and forging the metals together The following methods are classified under resistance welding (1) spot welding (2) seam welding (3) projection welding (4) flash welding (5) upset welding, and (6) *percussion welding* (Fig 1)

Most metals and their alloys can be joined to- [illegible] soft metals, such as copper silver and aluminum See WELDING AND CUTTING OF METALS

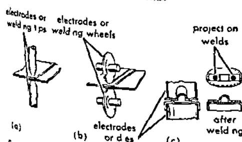


Fig 1 Three typ cal res stance-welding methods a) Spot weld (b) Seam weld (c) Project on weld (From A. L. Ph ll ps ed *Welding Handbook, 4th ed* American Weld ng Soc ety, 1957)

Power is usually obtained from the public utilities at the standard *frequency of 60 cycles per* second (cps), although one method converts 3-phase, 60-cps voltage into single-phase, 3–25 cps voltage for special *welding transformers* Welding current may also be supplied by a direct current or stored energy source

Alternating current is used *in about 90% of the* installations A current of several hundred amperes is used for thin metals, more than 100 000 amp may be used *for thick metals, depending on the* [illegible] volts, but the over all range for the various types of welding machines is 1–30 volts These low voltages *are obtained from a step-down transformer, which* usually has a cast copper, water cooled, single turn secondary

Direct current, when used, is obtained from various low voltage sources, such as rectifiers homopolar generators or storage batteries

Energy may be stored during a relatively long period and released suddenly from capacitors, magnetic fields storage batteries, or heavy flywheels on homopolar generators These types of power supply eliminate large transient loads on small power lines

Principles of operation. The welding current flows from the transformer through a flexible conductor the horn and electrode of the welder, the workpieces the other electrode and horn and back through a flexible conductor to the transformer Open circuit secondary voltage necessary to produce the welding current is determined by the impedance of the secondary circuit When this voltage is multiplied by the welding current the demand kilovolt amperes (kva) of the welding transformer is obtained Because the welding current *flows for a small part of the total time*, the demand kva rating is greater than the thermal rating

Total heat H generated in the workpieces and electrodes is expressed in watt seconds

$$H = I^2RT$$

where I is the current in amperes R is the sum of contact and workpiece resistances measured between the electrodes and T is the time in seconds

Mechanical structure Modern machines are designed with low inertia electrode systems so that welding electrodes maintain the proper forces on *the weld at all times If the electrodes fail to follow* through the increased contact resistance may cause excessive heating resulting in burning of the electrodes and explosions *in the overheated* weld metal Clamping force of the electrode is obtained by manual, mechanical, hydraulic and pneumatic means Electrodes should approach rapidly but in a controlled manner so they are not *hammered out of shape*

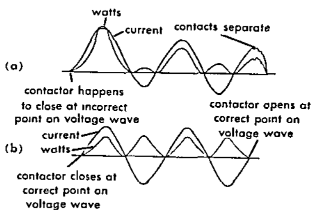


Fig 2 Curves show difference in current and power waves when circuit is closed nonsynchronously and synchronously (a) Closure of the circuit at the incorrect point with resultant decaying transient The opening is also nonsynchronous (b) Synchronous opening and closing of the circuit at the point of zero current corresponding to minimum transient (From A L Phillips, ed, Welding Handbook, 4th ed, American Welding Society, 1957)

Welding time controls. The length of time that the weld current flows can be controlled by manual, mechanical pneumatic or electronic systems Electronic controls exceed all other types in their ability to control precisely the length of weld time Times range from ½ cycle of a 60 cps voltage to several seconds depending on the thickness of the material In stored energy systems, such as capacitors or magnetic fields weld time is determined by the time constant of the electrical system

Current controls. The heating value of alternating current can be controlled electronically by synchronous phase shift heat controls, so that the point of current application on a sine wave will always be the same with respect to the voltage wave during the weld and for successive welds (Fig 2)

Current can be divided into pulses instead of a continuous flow This process is known as pulsation welding, when applied to spot or projection welding or as interrupted timing when applied to seam welding Various levels of pulses can be used to preheat the metal make the weld temper the quenched material and refine the grain in the weld

Electrodes must be water cooled to prevent their sticking to the workpieces The electrodes serve five purposes in the welding process (1) they apply the necessary force to bring the workpieces into intimate contact at the interfaces and into proper alignment, (2) they produce the proper contact resistances between the electrodes and the workpieces which include the interfaces, (3) they conduct the welding current, (4) they prevent spitting, porosity, and internal ingot cracking by maintaining the proper force on the weld, and (5) they dissipate the heat developed in the work pieces after the weld is made

The electrodes are made of copper alloys, which provide a combination of required electrical conductivity and mechanical strength *See* FLASH WELDING, SPOT WELDING [E.J.L.]

Bibliography A. L. Phillips (ed), *Welding Handbook*, 4th ed , 1957.

Resistivity, electrical

The electrical resistivity of a substance is the resistance which 1 cm³ of the substance offers to a flow of current when the flow is perpendicular to a pair of opposite forces of the cube The resistivity ρ, sometimes called specific resistance, is a property of the substance independent of its shape It is related to the resistance of a particular specimen of that substance by the relation

$$\rho = RA/l \qquad (1)$$

Here R is the dc resistance, A the cross sectional area normal to the current flow, and l the length of the specimen measured along the direction of current flow The unit of resistivity in the mks system is the ohm meter The resistivity is the reciprocal of the conductivity *See* ELECTRICAL CONDUCTIVITY OF METALS, RESISTANCE, ELECTRICAL

Variation with temperature. The resistivity of a particular substance is a function of temperature. For metals, except at very low temperature, ρ increases approximately linearly with absolute temperature For some metals, $\rho = \rho_0$ at absolute zero, where ρ_0 is called the residual resistivity The quantity ρ_0 is normally a small fraction of the resistivity at room temperature Other metals abruptly lose all of their resistance at a definite temperature (in all known cases below 20°K) and become superconductors *See* SUPERCONDUCTIVITY

The resistivity of semiconductors, such as silicon and germanium increases with falling temperature, and becomes exceedingly large as absolute zero is approached At low temperatures, semiconductors are essentially insulators

Finally there are some alloys such as constantan and manganin whose resistivities are nearly independent of temperature over a wide range To minimize temperature corrections, fixed resistors and coils in resistance boxes are generally made of such alloys

Variation with pressure. Resistivity is also a function of pressure and may either increase or decrease at high pressure, depending upon the substance In most cases the change is not appreciable until the pressure reaches hundreds of atmospheres or more This change with pressure arises for two reasons First under pressure the amplitude of vibration of the ions in the crystal diminishes, and consequently scattering of the conduction electrons by the lattice vibrations is reduced (*see* LATTICE VIBRATIONS) Second, the number of conduction electrons per unit volume increases with pressure

The effect of pressure on the resistivity of polyvalent metals is difficult to predict because changes in the band structure (*see* BAND THEORY OF SOLIDS) may be large and may lead to anomalously large

Pressure coefficient of resistivity $\alpha_I(\rho)$ of some metals at 1 atm

Metal	Temperature °K	$\alpha_P(\rho)$ 10^{-10} m²/kg
Lithium	303	−7 00
Sodium	303	58 8
Beryllium	298	1 77
Magnesium	298	5 4
Aluminum	301	4 29
Calcium	303	−9 48
Iron	303	2 42
Cobalt	297	0 96
Nickel	298	1 77
Palladium	299	2 10
Copper	303	1 92
Silver	303	3 48
Gold	303	3 02

(or small) and occasionally negative values of the pressure coefficient of resistivity $\alpha_P(\rho)$, which is defined by the relation

$$\alpha_P(\rho) = -\frac{1}{\rho}\frac{\partial \rho}{\partial P} \tag{2}$$

where P is the pressure. Some measured values of $\alpha_P(\rho)$ are listed in the accompanying table For additional information *see* HIGH PRESSURE PHYSICS

The resistivity of many substances is also dependent upon the state of *strain of the material* Reproducible measurements can only be made on carefully annealed samples

Typical values The usual method for determining resistivities is to measure the resistance of a wire of known length and cross section often by means of a Wheatstone bridge or similar circuit The resistivities in *ohm meters times* 10^{-8} *of some* common metals at 0°C are

Silver	1 50
Aluminum	2 50
Gold	2 04
Cesium	19 0
Copper	1 55
Potassium	6 3
Lithium	8 5
Magnesium	3 94
Sodium	4 27
Tungsten	4 89

Sources of resistivity The resistivity of metals arises from the scattering *of conduction electrons* by either lattice vibrations or stationary imperfections or both *In the* absence of all imperfections that is, in the ideal perfect crystalline lattice (which can never be attained) the resistivity of a metal would vanish. The phenomenon of superconductivity however is not due to the absence of lattice imperfections but arises from very different causes

The observed increase of resistivity with temperature is fundamentally a consequence of the increase in the thermal excitation of the lattice With increasing thermal agitation the probability of scattering a conduction electron out of its path is enhanced and consequently the resistivity increases

Conduction electrons may also be scattered by stationary imperfections Such imperfections *are* present in all metals or alloys, no matter how carefully they are prepared It is convenient to divide *these into three groups* point defects, cylindrical defects, and planar defects Point defects are imperfections which display approximate *spherical* symmetry in the lattice for example, impurities, vacancies, and interstitials. Cylindrical defects *are dislocations* Planar defects are stacking faults, grain boundaries, and the external surfaces of the specimen The resistivity due to stationary defects arises from elastic scattering of conduction electrons *See* CRYSTAL DEFECTS, DIFFUSION IN SOLIDS

Residual resistivity. It can be shown that

$$\rho = \rho_L + \rho_0 \tag{3}$$

provided lattice and impurity scattering are isotropic This equation known as Matthiessen's rule, states that the resistivity of a metal is the sum of the resistivity of the ideally pure metal ρ_L and *the resistivity due to imperfections* ρ_0 The ideal resistivity vanishes as the temperature approaches absolute zero It is for this reason that ρ_0 is called the residual resistivity The residual resistivity of a sample is a sensitive measure of its perfection *See* MATTHIESSEN'S RULE [F J B , J W ST]

der

urn

H

Wilson *The Theory of Metals*, 2d ed 1953

Resistor

A component of an electric circuit that produces heat while offering opposition or resistance to the flow of electric current All conductors exhibit resistance in varying degrees however the term resistor is generally used only to describe a device specifically used to *introduce resistance into an* electric circuit The unit of resistance measurement is the ohm Resistors are described by stating their

ance and stray shunt capacitance of the resistor

For a discussion of the property of resistance see

which they are used

Power resistors Such resistors range in size from about 5 watts to many kilowatts and may be cooled *by air convection air blast or water The smaller* sizes up to several hundred watts are used in both the power and electronics fields of engineering

Instrument resistors Direct current ammeters employ resistors as meter shunts to bypass the major portion of the current around the low current elements These high accuracy four terminal resistors are commonly designed to *provide a voltage* drop of 50 millivolts when a stated current passes *through the resistor*

Voltmeters of both the direct current and the alternating current types employ scale multiplying resistors designed for accuracy and stability The arc over voltage rating of these resistors is of importance in the case of high voltage voltmeters

Resistors for electronic circuits By far the greatest number of resistors manufactured are intended for use in the electronics field The majority of these resistors are intended for use in frequency selective circuits involving potentials up to several hundred volts but currents seldom over 10–100 milliamperes Their power dissipating ability is small, as is their physical size

Classification by construction. Resistors are also classified according to their construction, which may be composition, film type, or wire wound Further classification may be made according to whether the resistor has a fixed or adjustable resistance Adjustable resistors may be further classified as adjustable slide, rheostat, and potentiometer types

Composition resistor This resistor is by far the most widely manufactured type because of its low cost, reliability and small size Basically it is a mixture of resistive materials usually carbon, and a suitable binder molded into a cylinder Copper wire leads are attached to the ends of the cylinder and the entire resistor is molded into a plastic or ceramic jacket The over all length of the jacketed resistor excluding the leads is ⅓–1⅓ in for resistors varying in power rating from ¼ to 2 watts After manufacture the resistors are automatically zoned according to their individual resistance values which are indicated on the jacket of the resistor by means of a color code

Composition resistors are commonly used in the range from several ohms to 10–20 megohms and are available with tolerances of 20 10 or 5% Higher values of resistance are available but are not normally used in communication equipment or most electronic instruments For very high resistance values above about 100 megohms special jacketing is often required to prevent the leakage resistance over the surface of the jacket from altering the over all resistance of the resistor

All resistors possess a finite shunt capacitance across their terminals This capacitance is a function of the geometry and physical size of the resistor and is essentially independent of the value of the resistor The result is that at higher frequencies each resistor is effectively paralleled by a capacitive reactance which decreases in magnitude with increasing frequency At approximately 100 kilocycles the over all impedance of a 1 megohm resistor begins to decrease with an increase of frequency A 1000 ohm resistor will not display this effect until a frequency of about 100 megacycles is reached

The wattage rating of resistors is normally based upon the amount of thermal drift in resistance that can be tolerated If greater thermal stability is required the designer should use resistors with a higher power rating Composition resistors are easily damaged permanently by overheating Because of this, care must be exercised when soldering a resistor into a circuit

Film type resistor This resistor is rapidly replacing the composition resistor in applications where greater stability of resistance with voltage temperature, and humidity is demanded The design of the film resistor further lends itself to the controlled manufacture of precision resistors of any desired value Basically this resistor consists of a conducting film of carbon, metal or metal oxide deposited upon a ceramic cylinder The value of the resistance is controlled by controlling the thickness and length of the film The length of the film is often controlled by cutting a spiral groove around the resistor, the groove passing through the film to the insulating cylinder This spiral groove increases the effective length of the resistor and thereby determines its ohmic value By accurately controlling the pitch of the spiral the manufacturer can make a resistor of any value and maintain close manufacturing tolerances

The film resistor is often finished by coating it with an insulating varnish Often a plastic sleeve is slipped over the resistor to provide mechanical protection The spiral cut resistor displays a small inductive effect at the higher frequencies

Wire wound resistors Wire still remains the most stable form of resistance material available, therefore all high precision instruments rely upon wire wound resistors Wire-wound resistors are available in resistance ratings from a fraction of an ohm to several hundred thousand ohms, at power ratings from less than one watt to several thousand watts and at tolerances from 10 to 0 1% Because mechanical manufacturing problems limit the smallest wire size that can be used, these resistors are usually limited to values below about 100 kilohms Both inductive and noninductive types of resistor are manufactured

The inductive design is the common construction and consists of a spiral winding of wire about a cylindrical ceramic form After winding the entire resistor is covered with a vitreous material The spiral winding introduces a considerable amount of inductance into the circuit, which may become objectionable at the higher audio frequencies and all radio frequencies

The noninductive design includes several winding methods One of the simplest and most satisfactory is to reduce the cross sectional area of the coil by winding the wire around a thin flat card

Adjustable resistors The deposited film and wire-wound resistors lend themselves to the design of adjustable resistors or rheostats and potentiometers Adjustable slider power resistors are constructed in the same manner as any wire wound resistor on a cylindrical form except that when the vitreous outer coating is applied an uncovered strip is provided The resistance wire is exposed along this strip and a suitable slider contact can be used to adjust the over all resistance, or the slider can be used as the tap on a potentiometer

See POTENTIOMETER (VARIABLE RESISTOR), RHEOSTAT

Where continuous adjustment of the resistor is intended a ring shaped form is generally used For power resistors the ring is wound with resistance wire. For compact ½ and 1 watt resistors, the ring is coated on one surface with a resistance film Each type possesses all the advantages and disadvantages described above under fixed value resistors of its type In addition, adjustable resistors have the problem of maintaining a good noise-free, electrical contact at the wiper, which is mounted on a shaft concentric with the ring

For discussion of nonlinear resistors, see THERMISTOR VARISTOR. [R L.R]

Bibliography K. Henney, *Radio Engineering Handbook*, 1957, K Henney and C. Walsh (eds), *Electronic Components Handbook*, 1957, A E Knowlton (ed.) *Standard Handbook for Electrical Engineers* 1957, C. L. Wellard, *Resistance and Resistors*, 1968

Resolving power (optics)

A quantitative measure of the ability of an optical instrument to produce separate images The images to be resolved may differ in position because they represent (1) different points on the object, as in telescopes and microscopes or (2) images of the same object in light of two different wavelengths as in prism and grating spec

and for the latter class, as the smallest difference in wavelength or wave number that will produce separate ...

spectroscopes, however, for which the term chromatic resolving power is used signifying the ratio of the wavelength itself to the smallest wavelength interval resolved The figure quoted as the resolving power or resolving limit of an instrument is always the theoretical value that would be obtained if all optical parts were perfect. Aberrations of lenses or defects in the ruling of gratings will usually cause the actual resolution to fall below this value, and it therefore represents the maximum that could be obtained with the given ... of ... ixed ... ed ... theory *See* DIF

Chromatic resolving power The chromatic resolving power R of any spectroscopic instrument including prisms gratings and interferometers is defined as

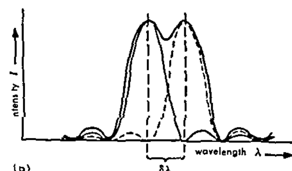
$$R = \frac{\lambda}{\delta\lambda}$$

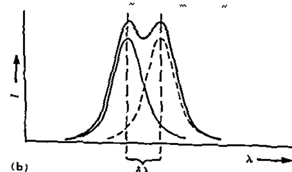
where $\delta\lambda$ represents the difference in wavelength of two equally strong spectrum lines that can barely be separated by the instrument, and λ the average wavelength of these two lines It is necessary to specify more precisely the term "barely separated," and for prisms and gratings in which the width of the lines is determined by diffraction, this is done by use of Rayleigh's criterion Part (*a*) of the figure shows the contours of two similar spectrum lines which are at the limit of resolution according to this criterion The lighter curve represents the line shape due to Fraunhofer diffraction for the wavelength λ, the dashed curve that for $\lambda + \delta\lambda$ and the heavy curve the sum of the two Rayleigh's criterion specifies that the lines are resolved when the principal maximum of one falls exactly on the first minimum (zero intensity) of the other Diffraction theory shows that the intensity I of either pattern at the central crossing point is $4/\pi^2$ of that at the maximum, so that the curve representing the sum dips to 81% at the center The theory also shows that the angular separation, $\delta\theta$ of the rays forming the two maxima is λ/a, where a is the linear width of the beam of light emerging from the prism or grating Hence, quite generally for such an instrument, the resolving power becomes

$$R = \frac{\lambda}{\delta\lambda} = \frac{\lambda}{\delta\theta}\frac{d\theta}{d\lambda} = a\frac{d\theta}{d\lambda}$$

Expressed in words, this result means that

$$\begin{pmatrix}\text{Chromatic}\\ \text{resolving power}\end{pmatrix} = \begin{pmatrix}\text{width of}\\ \text{emergent beam}\end{pmatrix} \times (\text{angular dispersion})$$





Resolution of two spectrum lines (a) when the shape is determined by diffraction (Rayleigh criterion) and (b) when the shape follows the Airy formula The latter is applicable to multiple-beam interferometers

In a given instrument the calculation of resolving power thus involves finding these two quantities

Resolving power of prisms When a prism is used at minimum deviation the resolving power depends on the length b of the base of the prism and the slope $dn/d\lambda$ of the dispersion curve giving the wavelength variation of the refractive index n Thus

$$R = b\frac{dn}{d\lambda}$$

Here the assumption is made that the prism is completely filled by the beam of light If it is not, b must represent the difference in path length between the longest and shortest rays through the prism See PRISM OPTICAL

Resolving power of gratings This equals the product of the order of interference m and the total number of rulings N The order m may be expressed in terms of the grating space s and the angles α and β of incidence and diffraction Thus

$$R = mN = \frac{Ns(\sin\alpha + \sin\beta)}{\lambda} = \frac{w(\sin\alpha + \sin\beta)}{\lambda}$$

where w is the width of the ruled area of the grating For the limiting case of grazing angles of incidence and diffraction the maximum possible R is seen to be $2w/\lambda$ or the number of wavelengths in twice the width of the grating See DIFFRACTION GRATING

Resolving power of interferometers For the type of interferometer most commonly used the Fabry Perot interferometer the resolving power may be expressed as the product of the order of interference $m = 2t/\lambda$ where t is the separation of the interferometer mirrors and an effective number N_{eff} of interfering beams For interferometers the line contour of the spectrum lines is not that of Fraunhofer diffraction but is given by a relation called the Airy formula This contour has no points of zero intensity but has the general shape shown in part (b) of the figure The Rayleigh criterion cannot therefore be applied in the usual way If however the two curves are made to cross at the half intensity point of each it is found that there is a dip of approximately 20% in the resultant curve The value of N_{eff} is thereby specified and the resolving power R is

$$R = mN_{eff} = m\left(\frac{\pi\sqrt{\rho}}{1-\rho}\right)$$

where ρ designates the reflectance of the interferometer plates See INTERFEROMETRY

Resolving power of telescopes This depends on the size of the diffraction maximum produced when light from a distant point source passes through a circular aperture of size equal to that of the objective lens or mirror A graph of the intensity in the diffraction pattern plotted against radial distance closely resembles one of the curves of the illustration (a) and hence the pattern consists of a central spot surrounded by faint rings The angular radius of the first dark ring corresponds by the Rayleigh criterion to the angular separation of two point sources that are barely resolved Theory gives this angle which represents the resolving limit as

$$\alpha = \frac{1\ 220\lambda}{d}\text{ radians} = \frac{14\ 1}{d}\text{ seconds of arc}$$

for λ = 5600 A and d the diameter of the objective lens in cm See TELESCOPE TELESCOPE, ASTRONOMICAL

Resolving power of microscopes This is determined by diffraction of a circular aperture representing the exit pupil of the microscope objective. There are two important differences in the resolving power of microscopes and telescopes. First, the resolving limit of microscopes is expressed in terms of the smallest distance l between two points on the object that are just resolved Second this limit depends on the mode of illumination of the object If the illumination is incoherent, so that there is no constant phase relation between light from adjacent points the resolving limit is

$$l = \frac{0\ 61\lambda}{n\sin\alpha}$$

where n is the refractive index of the material (for example oil) in the object space and α the angle that the extreme ray entering the objective makes with the axis of the instrument The quantity $n\sin\alpha$ is called the numerical aperture of the objective With coherent illumination the resolving limit is given by this formula with 1 0 in place of 0 61 provided the illumination is central When the object is illuminated from a point slightly to one side the factor may be reduced to 0 5 See MICROSCOPE OPTICAL SPECTROSCOPY [F A J]

Bibliography F A Jenkins and H E White, *Fundamentals of Optics* 3d ed 1957

Resonance (acoustics and mechanics)

When a mechanical or acoustical system is acted upon by an external periodic driving force whose frequency equals a natural free oscillation frequency of the system the amplitude of oscillation becomes large and the system is said to be in a state of resonance

When a simple oscillator of mass m stiffness constant s and mechanical damping constant R is driven by a periodic driving force $F\cos 2\pi ft$ it vibrates with a velocity amplitude

$$V = \frac{F}{[R^2 + (2\pi fm - s/2\pi f)^2]^{1/2}}$$

This equation implies that (1) the amplitude becomes a maximum when the driving frequency is $f = (1/2\pi)\sqrt{s/m}$ that is at the natural free oscillation frequency of the oscillator (2) small damping constants R are associated with large amplitudes of vibration at resonance and (3) the smaller R the more rapidly the amplitude de-

creases as the driving frequency departs from the resonance frequency In addition driving any vibrating system at its resonance frequency is characterized by a maximum dissipation of power

A knowledge of both the resonance frequency and the sharpness of resonance is essential to any discussion of driven vibrating systems When a vibrating system is sharply resonant, careful tuning is required to obtain the resonance condition Mechanical standards of frequency must be sharply resonant so that their peak response may easily be determined In other circumstances resonance is undesirable For example in the faithful recording and reproduction of musical sounds it either is necessary to have all vibrational resonances of the system outside the band of frequencies being reproduced or to employ heavily damped system See RESONATOR ACOUSTIC, SYMPATHETIC VIBRATION VIBRATION [L.E.K.]

Resonance (alternating current circuits)

A special condition of an alternating current circuit containing both inductance and capacitance Resonance is difficult to define in general terms since the resonant condition is different in series and parallel circuits The condition of resonance is of great importance in communications where certain frequencies may be either passed or rejected by a resonant circuit A few of the many applications of resonant circuits are filters radio receivers and transmitters oscillators and tuned amplifiers See ALTERNATING CURRENT CIRCUIT THEORY

Series resonance Resonance in a series circuit occurs when the inductive reactance equals the capacitive reactance and a condition of maximum current results The impedance of an RLC series circuit is

$$Z = \sqrt{R^2 + (X_L - X_C)^2}$$

where R is the resistance X_L is the inductive reactance and X_C is the capacitive reactance When $X_L = X_C$ the impedance Z will be equal to R and therefore will be a *minimum* The current will then have a maximum value of E/R, where E is the applied voltage

Inductive reactance and capacitive reactance are both dependent on the frequency f of the alternating current

$X_L = 2\pi fL$ where L is the inductance

$X_C = \dfrac{1}{2\pi fC}$ where C is the capacitance

Resonance occurs when

$$2\pi fL = \frac{1}{2\pi fC}$$

It can be seen that resonance can be obtained by varying frequency f inductance L or capacitance C. The frequency at which resonance occurs is called the resonant frequency f_0 If L and C are fixed values only one value of f_0 is possible From the solution of the preceding equation

$$f_0 = 1/2\pi\sqrt{LC}$$

At resonance the voltages across the inductance and capacitance will be equal but 180° out of phase and therefore cancel each other The value of these voltages will be Q times the voltage across the resistance where $Q = X_L/R$ See Q

Another condition of resonance occurs when the maximum energy stored in the electromagnetic field of the inductance is equal to the maximum energy stored in the electrostatic field of the capacitor

$$\tfrac{1}{2}LI_m^2 = \tfrac{1}{2}CE_m^2$$

where I_m and E_m are maximum values of the alternating current and voltage respectively

A typical resonance curve for the series RLC circuit is shown in Fig 1*a* which is a graph of the rms current as a function of frequency for a constant applied voltage

The pronounced effect of resistance particularly at and near the resonant frequency f_0 is indicated by the two curves one of which has twice the resistance of the other

The resonance curve (Fig 1*b*) is said to have a band width (BW) defined by

$$\mathrm{BW} = f_2 - f_1 = \frac{f_0}{Q_0}$$

where f_2 and f_1 are the frequencies at which the current is 70 7% of its peak value These points are called the half power points because the power

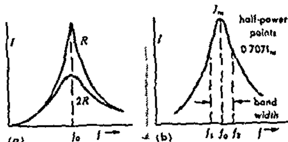


Fig 1 Series resonance—the *RLC* circuit (a) Typical resonance curve (b) Band width

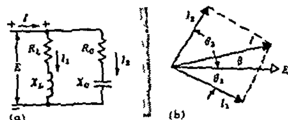


Fig 2 Parallel resonance (antiresonance) occurs when *I* and *E* are in phase (a) Parallel circuit (b) Current diagram

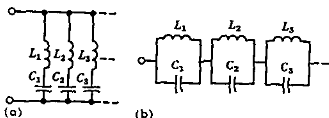


Fig 3 Examples of multiple resonance (a) Resonant paths in parallel (b) Resonant paths in series

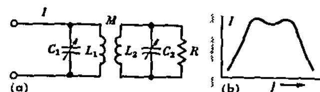


Fig 4 Resonance in the tuned rad o frequency circu t (a) Tuned rf circuit (b) Response curve

rical about f_0 However in the usual practical circuit where resistance is small Q_0 of the circuit is large and the sharpness of resonance is very pronounced Within the band width region the curve is nearly symmetrical and f_1 and f_2 are assumed to be separated from f_0 by the same frequency difference

Parallel resonance Resonance occurs in a parallel circuit when the out of phase components of the branch currents I_1 and I_2 are equal and opposite (Fig 2) For this condition to occur in the circuit of Fig 2 it is necessary that

$$\frac{X_L}{R^2 + X_L^2} = \frac{X_C}{R^2 + X_C^2}$$

If the resistances are small and can be neglected then the condition is that $X_L = X_C$ which is the same condition as for series resonance The resonant frequency is also the same as for series resonance

$$f_0 = 1/2\pi\sqrt{LC}$$

At resonance the current I of Fig 2 is in phase with the voltage E Parallel resonance is often called antiresonance to distinguish it from series resonance

There are many practical uses of the parallel resonance phenomenon Many circuits contain negligible resistance in the capacitive branch such as a capacitor in parallel with a coil

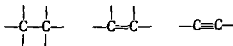
Multiple resonance Multiple resonance is the appearance of more than one resonant or antiresonant frequency in a circuit Figure 3 shows combinations of L and C that give rise to multiple resonance Each branch in Fig *3a* will show series resonance at a particular frequency and each branch of Fig *3b* will show parallel resonance at a particular frequency As frequency is varied each circuit also has resonance effects between branches or sections These circuits exhibit both series and parallel resonance and as frequency is increased from zero to a high value the current response of each will have a succession of alternating maximum and minimum magnitudes

An important application involving multiple resonance is the tuned circuit of Fig *4a* The input current response is shown in Fig *4b* The peaks resulting from the resonant frequencies of the two meshes can be moved closer together or farther apart by changing the degree of coupling between the meshes [B L R]

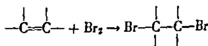
Resonance (molecular structure)

The term resonance is used in chemistry to express two concepts It may refer to a specific quantitative mathematical procedure called the valence bond resonance method for calculating the energy levels of electrons and certain other parameters of molecules but as commonly used it refers to the qualitative nonmathematical application of the method and is therefore equivalent to the term mesomerism In this sense resonance represents a refinement of structural theory which allows for nonintegral bonds The qualitative use of resonance is the subject of this article See MOLECULAR STRUCTURE AND SPECTRA

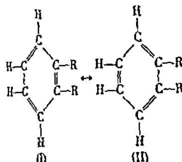
According to elementary structural theory the total number of bonds of an atom in a molecule is some small integer For a carbon atom this number is always four For example two carbon atoms may be joined together by 1 2 or 3 bonds such bonds are referred to respectively as single, double or triple bonds



Certain chemical reactions are associated with certain structural features Compounds which have a carbon-carbon double bond for example, are characterized by high reactivity They are readily oxidized by aqueous potassium permanganate solutions and undergo addition reactions with many reagents such as bromine

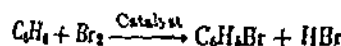


Resonance in benzene From the earliest development of structural theory (1858–1865) chemists have recognized that certain important compounds such as benzene do not fit into the simple theory Structurally benzene consists of a ring of 6 carbon atoms to each of which is attached a hydrogen atom (or in its derivatives some other atom or group such as chlorine OH CH_3 and so on) F A Kekule suggested that the structure be written as (I) or (II) (for benzene R = H) and in order to explain why there was only one known isomer of each ortho disubstituted benzene (R ≠ H) instead of the two predicted he further suggested that (I) and (II) are easily interconverted by movement of the double bonds The difficulty chemists found with this suggestion is that benzene does not have the properties commonly ob-



served with other compounds which contain double bonds. The discrepancies can be classed as chemical, structural, and thermochemical

Chemical discrepancies Chemically, benzene and its derivatives do not have the reactivity associated with simpler compounds containing double bonds. For example, benzene can be heated with permanganate solutions without change and with bromine, it reacts much less readily and does so by substitution rather than addition



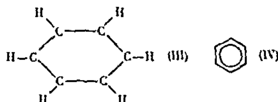
$$C_6H_6 + Br_2 \xrightarrow{\text{Catalyst}} C_6H_5Br + HBr$$

Structural discrepancies These became evident with the development of x ray diffraction electron diffraction and spectroscopic techniques for the measurement of interatomic distances The C—C single bond distance is about 1 54 A (1 54 × 10^{-8} cm) and the C═C double-bond distance is about 1.33 A. Thus, the Kekule structures (I) or (II) should have a ring with alternating long and short bonds, pulsating perhaps as the interconversion takes place X ray diffraction measurements on such benzene derivatives as hexamethylbenzene however show that all ring bonds are of equal length, an intermediate 1 39 A. Furthermore it can be concluded from the infrared and the Raman spectra that benzene has a sixfold axis of symmetry These measurements do not allow a Kekule structure, pulsating or not

Thermochemical discrepancies Finally thermochemical measurements of various types show further discrepancies Thus the heats of combustion of many compounds can be predicted quite accurately (± a few kilocalories out of about 1000 kcal) by adding up the contributions of the C—H bonds, the C—C bonds the C═C bonds and so on The predicted value for the heat of combustion of a Kekule benzene structure is about 35–40 kcal/mole greater ...

...nce concept of benzene structure For about 75 years, numerous and ingenious modifications of structural theory were proposed in an effort to accommodate the facts but it was only by the applying of quantum mechanical principles to electronic structures of molec... ...

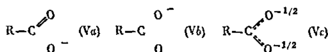
gen atom. *All of the atoms lie in the same plane,* and the molecule has hexagonal symmetry Between each hydrogen atom and the carbon atom to which it is attached is localized one pair of electrons (the valence bond electrons corresponding to the line of the Kekule structure or to a pair of dots of the electronic structure) Between each two carbon atoms there is also localized one pair of electrons These twelve pairs of electrons are called σ electrons It must be emphasized that it is necessary to picture such electron pairs as representing a diffuse region of high negative charge density between the atoms rather than as geometrical points The remaining six electrons, called the π electrons, are then assigned a special distribution in two doughnut shaped clouds, one on either side of the ring. *In the Kekule structures, these six* electrons make up the second bond of each of the three double bonds In effect then, each bond of the ring is to be represented by 1½ classical bonds Formula (III) shows the structure and (IV) is a useful schematic representation



In other words the actual benzene molecule is an intermediate between the two valence bond formulas (I) and (II) One way to describe this is to say that benzene is a resonance hybrid of structures (I) and (II) or to say that benzene exhibits resonance In the past this terminology has led *to some misunderstanding on two counts* It seems to imply that benzene is a rapidly interconverting mixture of structures such as (I) and (II), a conclusion at variance with the physical evidence (and not intended by resonance theory) Also, it does not provide a suitable term for describing the actual molecule The equivalent British term mesomerism is somewhat better in these respects Mesomerism is derived from roots meaning 'between the forms' thus implying that the structure is intermediate and it provides a term for referring to the actual molecule as the mesomeric molecule or *the mesomeric ion* *See* BENZENE

Resonance in other structures Whenever two or more classical (or electronic) structures can be written for one arrangement of the atoms, then none of these classical structures is a good representation of the molecule Instead the actual structure is intermediate This intermediate structure *will have an intermediate chemical reactivity,* intermediate bond lengths and it will have greater thermodynamic stability than predicted for any one of the classical structures

The following examples illustrate further the *use of the resonance concept* Structures (*Va*) and (*Vb*) are two electronic formulas that can be

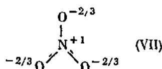


written for the known arrangement of the carbon and oxygen atoms of the carboxylate ion Neither structure is a good representation of the ion, instead it has some intermediate structure, as depicted in (Vc), in which each oxygen atom shares one half of a negative charge, and each carbon-oxygen bond is in effect 1½ bonds The physical and chemical properties are in accord with (Vc)

It is not necessary for the average to work out to any particular bond fraction, nor is the resonance concept limited to carbon compounds For the carbonate ion, there are three equivalent, although geometrically distinct, electronic arrangements that correspond to (VIa) When these are aver-



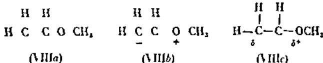
aged, the result is (VIb) in which each carbon oxygen bond is a 1⅓ bond and each oxygen atom has the same fractional charge (two-thirds of an electronic charge)



Carbonate ion can be referred to as a resonance hybrid of the three contributing structures corresponding to (VIa) or it can be referred to as a mesomeric anion

Nitrate ion is isoelectronic with carbonate To obtain its structure the symbol N is substituted for the symbol C and a formal positive charge is written on the N because nitrogen has one less valence electron than carbon, formula (VII)

In the compounds considered above all the contributing structures have been equivalent although geometrically distinct This condition is not neces



sary Two electronic structures can be written for methyl vinyl ether (VIIIa) and (VIIIb) Since (VIIIb) involves a separation of charge it represents a structure of higher energy than (VIIIa) As a consequence, the averaging process will weight (VIIIa) more heavily and the best that can be said qualitatively for (VIIIc) is that there will be some charge separation as shown and that the carbon carbon double bond is a little less than a double bond and the carbon oxygen single bond is a bit toward being a double bond These factors help to explain qualitatively the properties of methyl vinyl ether

Bond energies. There are two mathematical methods which can in principle be used to calculate quantitatively the energy levels and the electronic distribution in molecules The valence bond resonance method is a quantitative procedure based on the qualitative averaging described above The term resonance came to be used since the mathematical methods used are those which apply to resonating systems in general

The other mathematical method is the molecular orbital treatment which is more widely used today In this treatment, the system of atoms is considered as a framework onto which electrons can be fitted in much the same way that electrons are fitted into atomic orbitals to give atoms If carried through in sufficient detail, both the valence bond resonance method and the molecular orbital method give the same results *See* CHEMICAL BINDING, ORGANIC CHEMISTRY, QUANTUM CHEMISTRY

[D F D]

Bibliography G W Wheland, *Resonance in Organic Chemistry*, 1955

Resonance transformer

An electrostatic particle accelerator in which the high voltage terminal oscillates over the voltage range $\pm V$ These machines are used principally for acceleration of electrons Electron current is allowed to pass only when the terminal voltage is

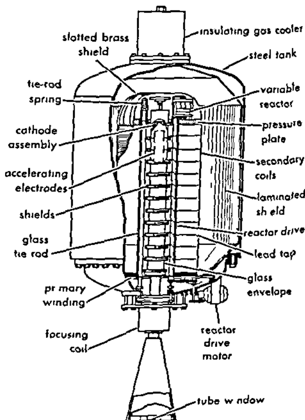


Fig 1 Schematic diagram of 1-Mev resonance transformer manufactured by the General Electric Company (General Electric Co)



Fig 2 A 2-Mev General Electric resonance transformer unit showing high voltage transformer stack (left) and steel tank (right) which encloses it (*General Electric Co.*)

near its peak value of $-V$ Thus the energy spread of the electrons can be held to a moderate value

The resonant transformer consists of a low voltage primary winding which surrounds the lower end of a high voltage coil stack The stack is made up of a number of thin flat windings called pancake coils, in which the multisection vacuum tube is coaxially mounted The inductance and capacitance of the high voltage secondary have values such that the resonant frequency of this circuit is equal to the frequency at which primary power is supplied These machines as manufactured by the General Electric Company utilize 180-cycle primary power The resonant transformer and accelerating tube are housed in a steel tank and high voltage insulation is provided by the use of sulfur hexafluoride gas. *See* PARTICLE ACCELERATOR

Resonance transformers are manufactured in three different sizes a 1 Mev 5 ma unit (Fig 1) a 2 Mev 6 ma unit (Fig 2) and a 3 5 Mev 8 ma unit They are used for industrial radiography for x ray therapy for food and drug sterilization and for the processing of plastics *See* COCKCROFT WALTON ACCELERATOR [R G H]

Bibliography *See* VAN DE GRAAFF GENERATOR

Resonator, acoustic

A device consisting of a combination of elements having mass and compliance whose acoustical reactances cancel at a given frequency Resonators are often used as a means of eliminating an undesirable frequency component in an acoustical system In other instances resonators are used to produce an increase in the sound pressure in an acoustic field at a particular frequency

Resonators are useful most often in the control of low frequency sound They are of particular value in reducing the noise from sources having constant frequency excitation For example a reciprocating air compressor driven at constant speed produces single frequency sound waves which can be attenuated most effectively by resonators

Resonators have also found considerable application in architectural acoustics It is often difficult to obtain adequate control of reverberation time at low frequencies in a large studio or auditorium using conventional acoustical materials A number of designs for these spaces have included the construction of resonators behind walls or in the ceiling to obtain increased low frequency absorption and thus provide more satisfactory reverberation characteristics Since the frequency range of efficient absorption is generally restricted for any given resonator many different units often with different resonant frequencies are employed In this way fairly broad frequency range absorption is obtained providing a relatively uniform reverberation time-frequency characteristic for the room

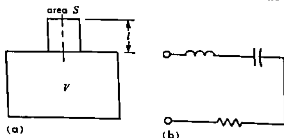
The Helmholtz resonator, illustrated below is the simplest and most often utilized acoustical resonator The unit consists of a straight tube of length l and cross-sectional area S, connected to a closed volume V This combination is directly analogous to the simple series LC electrical circuit

Resonant frequency In analogy to the electrical circuit the resonant frequency f_0 of the Helmholtz resonator is given by the equation

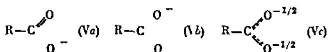
$$f_0 = \frac{1}{2\pi}\sqrt{\frac{c^2S}{l_eV}} \quad \text{cps}$$

where c is the speed of sound and l_e is the effective length of the tube approximately equal to $l + 0\,8\sqrt{S}$ The computation is normally valid if the linear dimensions of the tube are small compared to the wavelength of sound at the resonant frequency λ_0 This value is given by $\lambda_0 = c/f_0$ At higher frequencies the performance of the resonator must be computed from a distributed parameter or wave-analytical point of view

Damping In addition to the resonant design frequency the damping of acoustical waves by the resonator is of interest The damping may be described in terms of the quality factor Q in analogy to the quality factor in an electrical circuit (*see* Q) This factor describes the ratio of energy stored to the energy dissipated per cycle, in the resonator The ideal dissipationless resonator would have an infinite value of Q In reality viscous losses in the



(a) Helmholtz resonator (b) its electrical analog

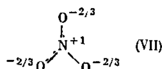


written for the known arrangement of the carbon and oxygen atoms of the carboxylate ion Neither structure is a good representation of the ion, in stead it has some intermediate structure, as depicted in (Vc), in which each oxygen atom shares one-half of a negative charge, and each carbon oxygen bond is in effect 1½ bonds The physical and chemical properties are in accord with (Vc)

It is not necessary for the average to work out to any particular bond fraction, nor is the resonance concept limited to carbon compounds For the car bonate ion, there are three equivalent, although geometrically distinct, electronic arrangements that correspond to (VIa) When these are aver



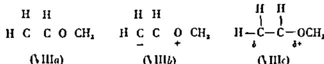
aged, the result is (VIb) in which each carbon oxygen bond is a 1⅓ bond, and each oxygen atom has the same fractional charge (two thirds of an electronic charge)



Carbonate ion can be referred to as a resonance hybrid of the three contributing structures corre sponding to (VIa), or it can be referred to as a mesomeric anion

Nitrate ion is isoelectronic with carbonate To obtain its structure the symbol N is substituted for the symbol C and a formal positive charge is written on the N, because nitrogen has one less valence electron than carbon, formula (VII)

In the compounds considered above, all the con tributing structures have been equivalent although geometrically distinct This condition is not neces



sary Two electronic structures can be written for methyl vinyl ether, (VIIIa) and (VIIIb) Since (VIIIb) involves a separation of charge it repre sents a structure of higher energy than (VIIIa) As a consequence, the averaging process will weight (VIIIa) more heavily and the best that can be said qualitatively for (VIIIc) is that there will be some charge separation as shown, and that the carbon carbon double bond is a little less than a double bond and the carbon-oxygen single bond is a bit toward being a double bond These factors help to explain qualitatively the properties of methyl vinyl ether

Bond energies. There are two mathematical methods which can in principle be used to calcu late quantitatively the energy levels and the elec tronic distribution in molecules. The valence bond resonance method is a quantitative procedure based on the qualitative averaging described above. The term resonance came to be used since the mathematical methods used are those which apply to resonating systems in general

The other mathematical method is the molecu lar orbital treatment which is more widely used to-day In this treatment, the system of atoms is con sidered as a framework onto which electrons can be fitted in much the same way that electrons are fitted into atomic orbitals to give atoms. If carried through in sufficient detail, both the valence bond resonance method and the molecular orbital method give the same results *See* CHEMICAL BINDING, ORGANIC CHEMISTRY, QUANTUM CHEMISTRY

[D F.D]

Bibliography G W Wheland, *Resonance in Organic Chemistry*, 1955

Resonance transformer

An electrostatic particle accelerator in which the high voltage terminal oscillates over the voltage range $\pm V$ These machines are used principally for acceleration of electrons Electron current is al lowed to pass only when the terminal voltage is

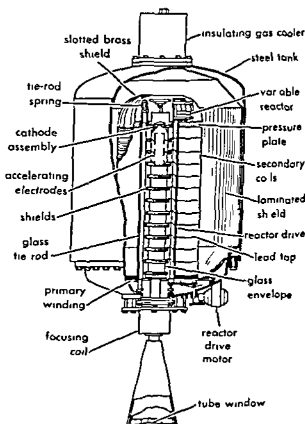


Fig 1 Schematic diagram of 1-Mev resonance transformer manufactured by the General Electric Company (General Electric Co)

Circulation and respiratory organs The most highly developed system for gas exchange consists of specialized respiratory organs in which oxygen and carbon dioxide are exchanged between the environment and blood and a circulation which carries this blood to the various body cells where the oxygen is given up and carbon dioxide absorbed. The respiratory organs expose the blood to the environment in vessels the capillaries so small in diameter that gas exchange with the environment can take place by diffusion alone in a very short time (Fig 2) The total surface area of these capillaries must be relatively large in order to permit the whole respiratory gas exchange of the animal to take place through them Such a large capillary surface area can be produced either by complex infolding of the body surface in which case the organs are called lungs or an outward extension of the blood vessels in which case the organs are called gills

Lungs These structures consist of a space lined with blood capillaries within the body of an animal The space may be a single cavity as in the case of simpler lungs or may have many millions of minute subdivisions the alveoli as in mammalian lungs (see LUNG) In mammalian lungs the small capillaries are of the order of 0 01 mm in diameter and the alveoli are of the order of 0 1 mm in diameter Lungs are usually air filled but structures like the respiratory tree may be water filled as in some aquatic animals, for example the sea cucumber (Fig 1f) In the case of air breathing animals lungs have the advantage over exposed gills of pre-

.

[illegible]ion lungs like those of scorpions Diffusion lungs are much less efficient than ventilation lungs and

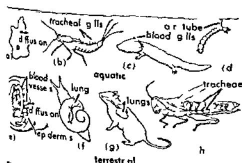


Fig 1 Types of respiratory mechanisms in animals (a) Protozoan, diffusion through cell wall (b) Mayfly nymph, tracheal gills (c) Salamander blood gills (d) Mosquito larva aquatic with tube for breathing free air (e) Earthworm diffusion through moist body wall to blood vessels (f) Land snail moist lung inside body (g) Land vertebrate pair of moist lungs inside body (h) Insect, tracheae throughout body (From T I Storer and R L Usinger General Zoology 3d ed McGraw-Hill 1957)

are confined to smaller animals Air may be moved in and out of ventilation lungs by positive pressure, as in the frog which forces air into the lungs by swallowing, or by expanding the space about the lungs as in mammals In man inhalation by the lungs is the result of the expansion of the bony framework of the chest through rib, muscular action and downward movement of the diaphragm the transverse sheet of muscle which separates the abdomen from the chest The breath is normally expelled by relaxation of these muscles

The transport of carbon dioxide in higher animals is accomplished by the reaction of carbon dioxide with chemical buffers in the blood a process in which hemoglobin is of primary importance (see HEMOGLOBIN) For the function of blood in the transport of oxygen see RESPIRATORY PIGMENTS

Gills Functionally the gills consist of fine blood vessels which are exposed externally to the environment (Fig 1c) They are generally confined to water dwelling animals because they need the mechanical support of the fluid However they may also be present in air breathers and in some animals are used both for air and water breathing Because gas diffusion is so slow in fluid it is essential that the water which envelopes the gills be stirred This is accomplished in the vertebrates by placing the gills so that they are exposed to the water through which the animal moves Because the efficiency of gas exchange depends in part on the rate of fluid movement past the gills it is necessary for some fish such as mackerel to keep swimming constantly to stay alive See SWIM BLADDER

Other special respiratory organs exist but are not of equal importance to lungs and gills Portions of the alimentary canal may be adapted for respiratory exchange for example the swim bladder in fish can store oxygen under some circumstances

Oxygen requirements Animal energy is derived from foodstuffs in a sequence of oxidation reduction steps the last of which is the reduction of molecular oxygen Although oxygen is the ultimate oxidant in this sequence animal life can be supported by the energy liberated in a sequence of these steps which does not require oxygen An aerobic animal is one that does require oxygen it completely oxidizes its foodstuffs and produces carbon dioxide Most animals are aerobic An anaerobic animal is one that does not require oxygen Most animals even those that are aerobic are able to derive at least part of their energy anaerobically for short periods of time or at different stages of development Under certain circumstances such as during activity the availability of oxygen in the cells of the aerobic animal may be insufficient to meet its energy demands the deficit can be met by anaerobic processes The difference is called an oxygen debt and must be repaid after restoration of normal conditions by aerobic reactions which dispose of the end products of the anaerobic reactions and replenish any depleted oxygen stores The oxygen debt after severe exercise can reach 11 liters in man

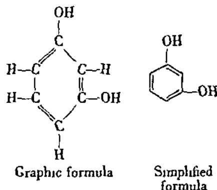
air passing through the neck of the resonator limit the value of Q to a maximum given by $Q_{max} \cong 270 f_0^{1/4}$

Energy absorption. The presence of dissipation tends to make the resonator tune over a narrow range of frequencies about the resonant frequency rather than solely at the resonant frequency itself In the case of a resonator used as an absorber, the value of Q defines the frequencies f_1 above and f_2 below the resonant frequency at which the efficiency of energy absorption is just one half that at the resonant frequency This relation is given by $Q = f_0/(f_1 - f_2)$

The maximum value of resonance energy absorption in a resonator, W_a, is given by $W_a = I\lambda_0^2/4\pi$ where I is the intensity of the incident acoustic wave The inherent dissipation in a resonator may be increased by placing an acoustical resistance, such as a porous cloth, across the opening of the neck In most applications this is a desirable feature because the resonance absorption can be maintained at a high value, but the frequency bandwidth over which the resonator will perform efficiently is increased See FILTER ACOUSTIC, IMPEDANCE, ACOUSTIC, MUFFLER [W J G]

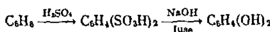
Resorcinol

A dihydric phenol in which the two hydroxyl groups are located meta to each other on the ring



Graphic formula Simplified formula

of carbon atoms It is produced by the disulfonation of benzene followed by fusion of the product with caustic soda



$$C_6H_6 \xrightarrow{H_2SO_4} C_6H_4(SO_3H)_2 \xrightarrow[\text{fuse}]{NaOH} C_6H_4(OH)_2$$

It reacts with formaldehyde to form resinous materials used as cold setting adhesives It is also an intermediate used in the manufacture of azo dyes and of fluorescein, a fluorescent compound from which dyes are made Resorcinol is also used to prepare hexylresorcinol, an antiseptic See PHENOL [R B C]

Respiration

Originally, this term meant breathing, but now it describes the over all process by which the individual cells of an animal are supplied with oxygen from and in turn give off carbon dioxide to the environment Oxygen is important as the ultimate oxidizing agent in the energy producing metabolic

dioxide are not secreted by animal cells, that is the gases are not made to move to a region of higher partial pressure by living processes, although there are thought to be some rare exceptions Respiratory exchange is therefore dependent upon diffusion and forced movement of either gas or fluid The rate of diffusion of a gas in a fluid per unit difference in gas partial pressure, is proportional to the solubility of the gas Because carbon dioxide is about 20 times as soluble in water as is oxygen, its rate of diffusion in water and in animal cells is 20 times as rapid Therefore it is usually the diffusion exchange of oxygen and not carbon dioxide which is the greater problem in respiration Because of its greater solubility, the environment of aquatic animals has a greater capacity for carbon dioxide than oxygen

Diffusion respiration. In the case of small organisms, less than about 1 mm in diameter, the distance from the surface to the center of the individual cells is so short (Fig 1a) that diffusion alone will exchange the gases with the environment at a rate sufficient to support their metabolism In larger animals however, extensive gas transport systems have developed to carry the gases over the relatively great distances from the body's surface to its cells The simplest improvement over diffusion alone is to conduct the environmental medium to the vicinity of the individual cells (Fig 1e) Some Coelenterata circulate ambient water through interior canals to accomplish this purpose Another outstanding example is the extensive system of narrow gas filled tubules or tracheae of the Arthropoda which supply fine branches to the body cells Because the diffusion coefficient of oxygen is about 1 000 000 times as great in gas as in water, these tracheae facilitate the respiration of cells within the animal In spite of this the rate of gas diffusion within the tracheae places a definite upper limit on the size of these animals Aquatic insects possessing air filled tracheae must seek the surface of the water intermittently to breathe Some carry bubbles down with them to act as an air store A number of aquatic insect larvae have developed closed gas filled tracheal systems Gas exchange in the tracheae with that in the water occurs through special respiratory surfaces tracheal gills analogous to the gills of fishes but filled with gas rather than blood (Fig 1b,h)

Diffusion and circulation. Another development in gas transport is that of a closed circulation in which a fluid is moved throughout the animal This convection aids the exchange of gas between the surface of the animal and its various component cells In spite of the absence of any special respiratory organs, sufficient gas exchange takes place between the circulating "blood" and the environment An example is the earthworm Even in some large animals, gas exchange across the integument is important

ssure (see CYTOCHROME) In addition there are mals whose oxygen consumption responds to anges in ambient oxygen pressure in a manner termediate between these two extremes As en ronmental oxygen pressure increases oxygen con umption increases until a certain oxygen pressure s reached above which it becomes relatively con tant. This oxygen partial pressure is called the critical tension An animal may respond to ex posure to an environment containing reduced oxy gen in several ways A completely aerobic animal may show a decrease in vital processes possibly cell death. An aerobic animal with some anaerobic processes may develop an oxygen debt A true anaerobe will obtain its energy from anaerobic processes indefinitely

Regulation of respiration In an aerobic animal respiration must be regulated ultimately by its rate of energy consumption The variables which control respiratory activity are oxygen tension and in higher animals carbon dioxide tension inside the animal's body because if respiration is inadequate compared to metabolic rate oxygen tension will fall and carbon dioxide tension rise

Vertebrates In the higher vertebrates rhythmic breathing movements are produced by nervous stimuli to the muscles of respiration originating from a region in the brain the medulla which is

volume of gas breathed per minute which tends to return carbon dioxide pressure of the blood toward normal This mechanism is very sensitive a rise in pressure of several millimeters of Hg will produce a doubling of volume rate of breathing in man. This mechanism is normally in control of

solved in the arterial blood acts upon chemically sensitive bodies, chemoreceptors located on the aorta and arteries in the neck to produce an increase in volume rate of breathing which tends to

stimulated by changes in lung distention and send back impulses to the respiratory center Although these impulses are able to modify the breathing movements, they are not intrinsically rhythmic and regulated breathing can exist in their absence al though the individual breaths may be considerably altered in form

The respiratory center is sensitive to temperature, an increase of which produces an increase in breathing. In addition increased body temperature increases the metabolic rate which will produce

Certain animals ...
mechanisms discussed above Diving mammals and birds are relatively insensitive to increases in arterial carbon dioxide tension in order to permit them to remain under the surface for long periods Some mammals have tremendously reduced respiration during hibernation largely as a result of the decrease in body temperature *See* HIBERNATION

The lower vertebrates have not been studied in as much detail but some type of respiratory center is generally present in the central nervous system The relative importance of oxygen lack as compared to carbon dioxide excess in the blood in regulating breathing has not been well established

Invertebrates In invertebrates a reduction in oxygen tension is a more important factor in controlling respiration than is carbon dioxide excess Regulation of respiration in these animals may take many forms The more primitive varieties avoid regions of low oxygen tension In mollusks the rate of pumping of water through the body increases at lower oxygen tensions In insects the orifices of the tracheae on the body surface can be opened and in some the mechanical ventilation of the tracheae increases when oxygen tension in the environment falls [R E F]

Bibliography A Krogh *The Comparative Physiology of Respiratory Mechanisms* 1941, C. L Prosser (ed) *Comparative Animal Physiology* 1950

Respiration, external

The processes by which oxygen (O_2) is carried into living cells from the outside environment and by which carbon dioxide (CO_2) is carried in the reverse direction In the simplest organisms this exchange is accomplished by diffusion of these gases directly between the cell and the outer environment to which it is exposed. In higher animals the route is more devious In man and other mammals breathing moves gas between the outer environment and the alveoli of the lungs The gas in the alveoli is separated from the blood in the pulmonary capillaries by a thin membrane Transfer of gases between alveoli and blood occurs across this membrane by diffusion of gases in physical solution

Diffusing capacity of lung The ability of the

transferred per unit time per unit of mean pressure gradient for the particular gas between alveolar gas and pulmonary capillary blood The diffusion process is normally so effective that gases in the alveoli are almost in equilibrium with the gases in the blood leaving the lungs that is the partial pressures of a given gas in the two phases are nearly identical *See* LUNG

Since there is no appreciable exchange of gases along the systemic vascular tree until the afferent end of the tissue capillaries is reached the partial

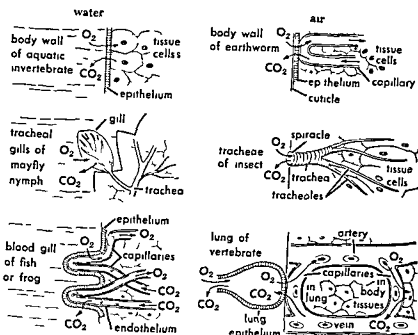


Fig 2 Equivalent nature of various respiratory mechanisms in different animals living in water or air, diagrammatic. (*From T I Storer and R L. Usinger, General Zoology, 3d ed, McGraw-Hill, 1957*)

Oxygen consumption. The number of milliliters of pure oxygen consumed by an animal per minute is known as the oxygen consumption In aerobic animals it is a measure of the metabolic rate because the source of their energy lies in the reaction of oxygen with the foodstuffs The energy produced by the consumption of 1 liter of oxygen varies with the type of food consumed, being 5 1, 4 5, and 4 8 kcal/liter for carbohydrate, protein, and fat, respectively Because the oxygen consumption of an animal depends on its metabolic rate it will vary with the factors that influence the latter such as activity body size and temperature, nutritional state, species, season, and sex (*see* METABOLISM)

Respiratory quotient is the ratio, rate of carbon dioxide production/rate of oxygen consumption Its importance lies in the information it gives concerning the type of food being oxidized in the animal The respiratory quotient has a value of 1 0 when carbohydrate is burned 0 79 when protein is burned, and 0 71 when fat is burned The normal respiratory quotient in man is about 0 80

Availability of oxygen. Because oxygen enters the animal's body by diffusion the partial pressure of oxygen in the environment must be greater than that in the animal at the point of entry Only that part of the oxygen in the environment which is at a partial pressure greater than that in the animal can be absorbed The atmosphere contains about 21% oxygen, corresponding to a partial pressure of 160 mm Hg in dry air at sea level Although air at higher altitudes also contains 21% oxygen the total pressure is greatly reduced and animal life becomes limited Aquatic animals depend upon oxygen dissolved in the water However, oxygen is relatively insoluble so that, whereas at sea level 5 ml of air contains approximately 1 ml of oxygen at standard conditions, it takes 154 ml of water equilibrated with air at 20°C to contain an equivalent amount of oxygen Because the oxygen in any body of water comes from the air originally, if the deeper waters are not mixed with the surface waters, their oxygen content may be decreased below that necessary to support life, particularly if any oxidative processes are going on such as decomposition of vegetation *See* FRESH WATER ECOSYSTEM

Elimination of carbon dioxide also occurs by diffusion and depends on the tension of the gas in the animal being greater than that in the environment, which is approximately zero under most circumstances Some natural waters containing carbonates may have considerable amounts of carbon dioxide in solution

Oxygen utilization is the percentage of the total oxygen in the air or water passing over the respiratory surfaces that is absorbed Some clams have an oxygen utilization of several per cent, that in resting man is 21% and the normal value in trout approximates 80% Although the oxygen utilization is to some extent a measure of the efficiency of the respiratory mechanisms it also depends on the partial pressures of oxygen in the animal and in the environment and upon the rate of air or water flow over the respiratory surfaces

Oxygen partial pressure and consumption Some animals show an increased oxygen consumption with increasing ambient oxygen tension However, many others such as the warm blooded animals are able to maintain approximately constant oxygen consumption over the wide range of environmental oxygen tensions which are compatible with life This adaptability depends in part on control of respiratory movements which try to compensate for the environmental lack in part on the efficiency of the respiratory pigments in transporting oxygen, and in part on the characteristics of the respiratory enzyme system whose rate of oxygen reduction is relatively independent of oxygen

. . to the environment

culatory that is they aid in oxygen transport These release their oxygen to the tissues at relatively high pressures Others appear to be primarily storage or intracellular transfer hemoglobins, these have extremely high oxygen affinities Myoglobins belong to the second category, they pick up oxygen released by the blood pigment and facilitate the transport of oxygen from muscle cell surface to the respiratory enzymes of the mitochondria which consume the oxygen Some animals possess hemoglobins for which a function is not apparent. Some of these animals survive for weeks even though their hemoglobin is rendered incapable

CHONDRIA

Microorganisms Among microorganisms hemoglobin is found in certain protozoa (*Paramecium* and *Tetrahymena*), in yeast and the mold *Neurospora* The root nodules which fix molecular nitrogen in leguminous plants are pink as a result of the presence of hemoglobin Both the fixation process and the occurrence of hemoglobin require the presence of the symbiotic bacterium *Rhizobium* Because hemoglobin is present only during periods of active fixation it is possible that hemoglobin might play some role in the fixation process These facts suggest the possibility that there may be functions of hemoglobin other than the transport or storage of oxygen They might function as oxida

cellular animals Certain bacteria possess an enzyme, hydrogenase which is concerned with the metabolism of molecular hydrogen This enzyme may be related to respiratory pigments because it combines reversibly with oxygen and contains iron However the presence or absence of heme or an other porphyrin has not been conclusively determined

Multicellular organisms Among multicellular invertebrates hemoglobins are distributed in every major phylum except those of the sponges and coelenterates. The pigment occurs in the parenchyma of the flatworm *Phaenocora* and in the blood cells of a few nemertean worms Hemoglobins occur in several parasitic nematodes including

...us, it takes about 150 seconds to deoxygenate the hemoglobin of the perienteric fluid of *Ascaris*, under the same conditions the dissociation time for sheep hemoglobin is 0 008 seconds *Strongylus* hemoglobin binds oxygen so tightly that the worms die of anoxia when placed in an anaerobic environment before their hemoglobin loses its oxygen This fact suggests that the hemoglobin may not be functional However, the hemoglobin in the body wall of *Ascaris* may be functional, for all body motion stops when the deoxygenation of the pigment starts

Among mollusks hemoglobin is found chiefly in the gastropods but has been found among a few lamellibranchs It also occurs in the muscles of those mollusks whose blood pigment is hemocyanin The whelk *Busycon*, has muscles with a greater myoglobin content than dog heart muscle Hemoglobins are found in many annelid worms In those worms with a closed circulation the hemoglobin is usually dissolved in the plasma, in those with less well developed circulatory systems the hemoglobin is usually in cells confined to the coelomic fluid In two, *Terebella* and *Travisia* hemoglobin apparently occurs both in the plasma and in the cells of the coelomic fluid Hemoglobins occur in the blood plasma of several crustaceans, including certain species of the water flea *Daphnia* Its presence in *Daphnia* depends on the oxygen content of the water, the lower the oxygen content the greater is the quantity of hemoglobin synthesized The ovaries and parthenogenetic eggs of these *Daphnia* contain a high concentration of hemoglobin which may help the eggs survive periods which are low in oxygen The sea cucumber *Thyone*, an echinoderm, contains a hemoglobin in the cells Hemoglobin has been found in only a few adult insects, all of which are aquatic and in two larval insects, each of which develops in an aquatic low oxygen environment The tracheal tubes of the backswimmers, *Buenoa* and *Anisops* are surrounded by clusters of cells which contain hemoglobin It occurs in the tracheal cells of the larva of the bot fly, *Gastrophilus*, which is found in horse stomachs It also occurs in the waterboatman *Macrocorixa* and in the body fluid of the larva of the midge, *Chironomus* No hemoglobin has yet been discovered in

are usually ular weights else they are extracellular and very large with molecular weights ranging from 400 000 to 3 000 000 For example a hemoglobin with a molecular weight of 3 000 000 occurs in the plasma of the lugworm *Arenicola* However another polychaete *Notomastus*, has a cellular hemoglobin of molecular weight 32,000 The larger the molecule the greater is the number of heme prosthetic groups which are bound to it Thus *Notomastus* hemoglobin has 2 hemes whereas *Arenicola* hemoglobin has about 180 Perhaps the confining of the small molecules in cells is a device which prevents their loss It is also possible that the large hemoglobin molecules in free solution in the body fluids of various marine animals helps them to maintain an osmotic balance of salts.

pressures of O_2 and CO_2 in the blood entering these capillaries depend mainly on the composition of alveolar gas The latter in turn is dependent on (1) the rate at which gas is transferred between lung and blood, (2) the rate at which the alveoli are ventilated, and (3) the composition and barometric pressure of gas with which the lung is ventilated If the composition of inspired gas is kept constant, the partial pressures of CO_2 and O_2 in the alveoli depend on the ratio of CO_2 output to alveolar ventilation and of O_2 uptake to alveolar ventilation respectively

Alveolar ventilation. By various feedback mechanisms the activity of the muscles of breathing and hence the alveolar ventilation is so precisely regulated that the partial pressures of these gases in the alveoli and, consequently, in the blood entering the tissue capillaries are kept relatively constant

The alveolar ventilation is always less than the total ventilation because only a part of each breath, called the tidal volume, actually reaches the alveoli Some of it is left, at the end of inspiration, in the conducting airways in which no gas exchange with the blood occurs and from which it is displaced to the outside during the ensuing expiration This part of each breath is called the dead space volume Thus, alveolar ventilation is equal to total ventilation minus dead space ventilation

In carrying out the breathing movements that result in pulmonary ventilation, the muscles of breathing are helped or hindered by several types of forces The chest and lungs are elastic in nature and must be stretched as the volume increases during inspiration The energy used for this purpose is stored and is available for accomplishing, or assisting, expiration Energy is also required during both inspiration and expiration to overcome the viscous resistance of gas flow through the tubes composing the respiratory tract and to produce plastic deformation and sliding movements of tissues Energy so used is not recoverable but is dissipated as heat Since the system is almost constantly being accelerated and decelerated inertia is also a factor but under most circumstances it is negligible

The rate of flow of gas in or out of the alveoli at any moment depends on the magnitude of the pressure gradient between the alveoli and the outer environment (alveolar pressure) and on the resistance to flow offered by the conducting airways The volume to which the lung is distended at any moment depends on the magnitude of the pressure gradient between the inside of the alveoli and the outer surface of the lung (transpulmonary pressure) and on the compliance of the lung The sum of the alveolar pressure and the transpulmonary pressure is called the intrapleural pressure

In the blood, O_2 and CO_2 are carried partly in physical solution and partly in readily reversible chemical combinations Exchange of gases between capillary blood perfusing the tissues of the body and the cells composing these tissues occurs by diffusion For this exchange to occur at an adequate rate, partial pressures of the gases in capillary blood must be maintained at levels such that the appropriate diffusion gradients are maintained Factors determining the partial pressure of a gas in a tissue capillary are the concentration of the gas in the blood entering the tissue capillary, the rate at which the gas is produced or consumed by the tissue cells, the rate of blood flow through the tissue, and the chemical affinity and solubility of the gas in the blood *See* RESPIRATORY SYSTEM

[A B O]

Bibliography: J H Comroe et al, *The Lung*, 1955

Respiratory pigments

Colored, metal containing proteins which combine reversibly with oxygen, found in the body fluids or tissues of invertebrate animals These pigments aid in the transport and temporary storage of molecular oxygen Thus they are distinguished from respiratory enzymes which are concerned with the metabolic consumption of oxygen Four distinctly colored groups of respiratory pigment exist among invertebrates hemoglobins (purple, become orange red with oxygen), chlorocruorins (green, become red when concentrated), hemerythrins (colorless, become red with oxygen), and hemocyanins (colorless, become blue with oxygen) Formerly invertebrate hemoglobins were called erythrocruorins to distinguish them from similar yet distinct pigments of vertebrate bloods Those hemoglobins confined to muscle cells are called myoglobins *See* HEMOGLOBIN

Composition. Each of the pigments is composed of two parts, a large protein molecule to which is bound one or more small moieties called prosthetic groups each of which contains a metal It is the metal which binds the oxygen, and this binding imparts the characteristic color to the pigment In hemoglobins the prosthetic group is an iron porphyrin compound called heme Chlorocruorin contains a similar iron porphyrin which differs only in that a vinyl group in the molecule is replaced by formyl The prosthetic group of hemerythrin consists of two adjacent iron atoms which bind an oxygen molecule between them, a third iron atom is present but it does not participate in the oxygen binding The prosthetic group of hemocyanin is analogous and consists of two adjacent copper atoms Pigments containing vanadium have been found in tunicates but these substances do not combine reversibly with oxygen and so cannot be considered respiratory pigments

The protein part of the pigment confers reversibility upon the combination of the metal with oxygen In the absence of protein, the prosthetic groups lose their capacity to combine with oxygen reversibly Instead the metals are irreversibly oxidized In hemoglobin the iron remains ferrous when combined with oxygen For this reason the combination of hemoglobin with oxygen is described as an oxygenation rather than an oxidation The protein is also responsible for certain physiological

Hemerythrin Hemerythrin is similarly structed and is found in a few annelid and gephyrean worms and the brachiopod *Lingula* It is found only in cells and has a molecular weight of 68,000 As in hemocyanin the oxygen is bound between two metal atoms each of which appears to be linked to sulfur atoms in the protein The difference is that the metal is iron in hemerythrin and copper in hemocyanin [A F R]

Bibliography D L Fox *Animal Biochromes and Structural Colours* 1953 S W Fox and J F Foster *Introduction to Protein Chemistry* 1957 F Haurowitz, *Chemistry and Biology of Proteins,* 1950 C L Prosser (ed), *Comparative Animal Physiology* 1950

Respiratory system

The system of organs involved in the acquisition of oxygen by an organism The lungs and gills are the two most important structures involved in the phase known as external respiration in which gaseous exchanges occur between the blood and environment. Internal respiration refers to the gaseous exchanges which occur between the blood and cells Certain other structures in some species of verte

...

fans is important in respiration Certain species of fish have a vascular rectum which is utilized as a respiratory structure water being taken in and ejected regularly by the animal Saclike cloacal structures occur in some aquatic species of turtles These are vascular and are intermittently filled with and emptied of water It is thought that they may function in respiration During embryonic life the yolk sac and allantois are important respiratory organs in certain vertebrates

Structurally respiratory organs usually present a vascular surface that is sufficiently extensive to provide an adequate area of absorption for gaseous exchange This absorptive surface is usually moist and thin enough to allow for the passage of gases This article treats the embryology and physiology of the gills and embryology anatomy and histology of the vertebrate lung and vertebrate respiratory physiology *See* ALLANTOIS GILL (ANATOMY), HEMOGLOBIN LUNG, RESPIRATION EXTERNAL RESPIRATORY PIGMENTS YOLK SAC

GILL DEVELOPMENT

...

Rhabdopleura, there is a pharyngeal region in which there are one or more (usually many) paired visceral clefts with delicate intervening strands of tissue which constitute the primary *bars or arches*, these are greatly augmented in some cases by the down growth of tongues of tissue between them which are termed secondary bars There are also transverse connecting strands termed synapticulae in some species such as *Amphioxus* The bars and synapticulae usually contain supporting gelatinous or chitinoid rods, in *Amphioxus* the primary bars contain prolongations of the dorsal coelom Most important, blood circulates through the bars and synapticulae during respiration and is aerated by water drawn in through the oral opening by the ciliated endoderm and passed out through the interstices of the basketlike structure This arrangement also acts as a strainer to retain food particles, a situation duplicated by the animals with gills (Fig 1)

The gills of chordates are located in the pharyngeal region They are always related to the visceral arches and clefts The latter structures are so defined because in vertebrates that respire entirely with lungs the arches and clefts or incipient clefts, lack gills which exist only in the embryo and disappear or are highly modified in the adult They always arise by the outpushing of pouches of endoderm which are usually dorsoventrally elongated through the mesoderm which contact corresponding inpushings of ectoderm Where these meet and break through clefts are formed and the intervening concentrations of mesoderm form the arches which are covered internally by endoderm and ... ectoderm The number of clefts the... wh... wi...

Visceral arches in craniates In this group the typical number of paired visceral arches and pouches or clefts is six including the mandibular and hyoid arches with the intervening spiracle However the number may be greater in some elasmobranchs and cyclostomes and less in some of the animals where these structures occur only in the embryos In these cases the number of pouches which actually become clefts is usually very limited for example there are three pairs in the

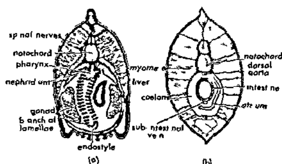


Fig 1 Amphioxus lanceolatus Transverse section of (a) pharyngeal region (b) intestinal region (From T J Parker and W A Haswell A Textbook of Zoology, 6th ed Macmillan 1956)

Oxygen carrying capacity The oxygen carrying capacity of bloods is greatly increased if they contain hemoglobin In the absence of any pigment the maximum amount of oxygen which 100 ml of blood can physically dissolve is only about 0 3–0 5 ml The blood of *Arenicola* has a capacity of up to 10 ml of oxygen/100 ml of blood Thus the hemoglobin accounts for at least 95% of the oxygen carried by the blood of this animal If a blood pigment is to be efficient as a transport agent it must not only bind oxygen but must also be capable of releasing it at sufficiently high pressures to meet the needs of the animal The higher the oxygen pressure at which the pigment unloads its oxygen the better will be the supply of oxygen in the tissue capillaries The more active a particular species of animal the greater in general is the oxygen pressure at which its blood unloads its oxygen An increased temperature lowers the affinity with which all blood pigments bind oxygen Thus an increased temperature not only accelerates metabolism generally but also tends to make more oxygen available by raising the pressure at which the hemoglobin delivers oxygen to the tissues provided that the hemoglobin is capable of becoming fully saturated with oxygen at the pressures existing in the region where oxygenation occurs

Many hemoglobins aid in the transport not only of oxygen but also of carbon dioxide Vertebrate blood hemoglobins have a lower oxygen affinity in the presence of carbon dioxide than in its absence This phenomenon the so called Bohr effect is also found in many although not all invertebrate blood pigments The effect facilitates the unloading of oxygen in the tissue capillaries and the discharge of carbon dioxide at the gills or other gas exchanging surfaces A pronounced Bohr effect is absent from most high affinity storage hemoglobins

The distribution of hemoglobins among invertebrate animals does not fit any phylogenetic scheme It appears as if hemoglobins have arisen independently many times in the course of evolution In some species the hemoglobins appear to be essential for the animal s survival in others the advantages appear marginal at best and perhaps their occurrence is fortuitous This may be partially explained by the fact that closely related compounds (the respiratory enzymes) are found in all animals and that the heme prosthetic group is present in every aerobic organism

Hemocyanins Hemocyanins are found only in mollusks and arthropods other than insects However they are not the only copper proteins which combine reversibly with oxygen The enzyme ascorbic acid oxidase found in plants has about the same copper content In the presence of oxygen this enzyme is blue as is hemocyanin and when the oxygen is removed the pigment becomes colorless exactly as does hemocyanin Hemocyanin itself has no oxidase activity Hemocyanins always occur in the plasma never in cells The molecular weights are as high as 7 000 000 The pigment is found in many of the higher crustaceans such as crabs and lobsters in the horseshoe crab *Limulus* and in scorpions Among the mollusks it occurs in both the cephalopods (squid and octopus) and in certain gastropods The squid is an active carnivore and requires well aerated sea water for survival Both the squid and the octopus possess bloods with about the same oxygen capacity yet the two hemocyanins have very different properties The hemocyanin of the squid is half saturated with oxygen at a pressure of about 36 mm Hg whereas octopus hemocyanin is half saturated at an oxygen pressure of only 3 mm Hg This large difference reflects the fact that the squid is a more active animal than the octopus About 92% of the oxygen of the arterial blood of the squid is normally removed during circulation through the tissues in man no more than about 30% is so removed Thus the squid is poorly adapted to survive even very short periods in an environment with little oxygen *Limulus* on the other hand survives weeks with all of its hemocyanin removed

The study of hemocyanin has had a curious history Most of the people who have studied it had previously studied hemoglobin and many of the properties of hemoglobin were unconsciously assumed also to hold for hemocyanin Because hemoglobin combines with carbon monoxide it was assumed that hemocyanin would do likewise The fact that the color and optical absorption spectrum of the supposed carbon monoxide hemocyanin compound was identical with that of hemocyanin without carbon monoxide was ignored Hemoglobin may be oxidized to form a compound called methemoglobin oxygenated hemocyanin was assumed to form a green methemocyanin upon reaction with the oxidizing agent ferricyanide which is yellow The fact that the mixture of any blue and yellow pigments will produce a green material was ignored However treatment of deoxygenated hemocyanin with one equivalent of hydrogen peroxide per copper atom results in a product called methemocyanin which lacks the ability to bind oxygen Hydrogen peroxide has no such effect on oxygenated hemoglobin

The unconscious attempt to fit hemocyanin into the hemoglobin mold thus helped to delay for at least 20 years the discovery that the oxygen of hemocyanin is bound between two copper atoms which are initially reduced (cuprous) The blue color of hemocyanin appears to be the result of the reversible transfer of an electron from at least one of the cuprous atoms of the oxygen thus at least one of the cuprous atoms becomes cupric Studies of the displacement of the copper by mercury compounds indicate that each copper atom must be bound to sulfur in the protein

Chlorocruorin Chlorocruorin has properties very similar to hemoglobin It is restricted to certain sessile marine annelids One genus *Spirorbis* contains several species of which one contains chlorocruorin in its blood another has hemoglobin and a third has no pigment at all Another closely related worm *Serpula* has both chlorocruorin and

tissue and a little muscle and covered where it is free, with epithelium The anterior and posterior layers on each side of the septum then become half gills or hemibranchs and the whole structure a holobranch (Fig 3)

Along the surface of each anterior and posterior layer there arises a series of transverse folds of tissue extending outward from the arch (Fig 4) These are primary lamellae covered by an epithelium, flat for the most part but with some columnar or cuboidal secretory cells Between the epithelial cells there is connective tissue which contains throughout each lamella a loop from the afferent and adjacent efferent vessel of the arch Another series of folds transverse to these lamellae develops next on the dorsal and ventral surface of each primary lamella They are the secondary lamellae whose internal surfaces are connected by numerous columnar cells which send out delicate protoplas lds of fer ent and efferent loops in the respective lamella (Fig 5) A cartilaginous ray extends outward along the base of each primary lamella at the boundary between the posterior layer or hemi

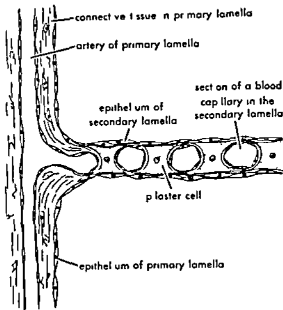


Fig 5 Diagram of a sect on through a part of a secondary lamella showing connection w th a primary lamella (*Mod fed after Goodrich*)

branch and the septum and helps to support the entire gill (Figs 3 and 4c) In the elasmobranch gill the young primary lamellae of the posterior hemibranch grow outward as filaments beyond the edge of the special type of elasmobranch operculum (Fig 3) These temporary filaments are therefore essentially external gills although in both origin and structure they are unlike the 'true external gills described below Because they occur only in the unhatched fish where they float in the albuminous fluid of the egg, they are probably as much for food absorption as respiration The septum of each holobranch grows out beyond the gill in these fish and turns posteriorly beneath the outer skin to the posterior edge of the following cleft, thus forming for the latter an individual cover or operculum Each branchial arch develops an anterior and posterior row of papillae on its internal border sometimes covered with enamel These are the gill rakers They act as strainers to prevent the escape of food (Fig 3)

Gills of Holocephali In this subclass the gills are similar to those in the elasmobranchs although arranged slightly differently with respect to the arches

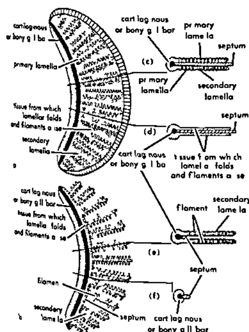


Fig 4 Diagrammat c representat ons (a) Poster or view of gill arch and g ll (hem b anch) from right side of the head of an elasmobranch (b) Same view of teleost arch and g ll (c) Hor zontal sect on along the leng h of a pa r of pr mary lamellae and the septum of the holobranch of an elasmobranch (d) Horizontal sec ion through three t ssue layers between primary lamel lae of an elasmobranch (e) Horizontal sect on along the leng h of a pa r of f laments of a teleost holobranch. (f) Hor zontal section between the f laments of a teleost holobranch hence through the arch and remains of the septum only

Cyclostome gill Although this class may be considered to be the most primitive of all the craniates its consideration has been delayed because the gills present a highly specialized condition most easily described as a variant of the elasmobranch and holocephalic type The beginning development is similar to that in the two former classes The septa are soon drawn out however so that the clefts become elongated tubes that sometimes open separately as in *Bdellostoma* or unite to open through a single orifice as in *Myxine* The gills proper have primary and secondary lamellae

chick, one pair in the cow, and normally none in the pig and man, in which the number of arches is reduced to five with the last vestigial The blood vessels found in the arches in animals other than fish are of considerable phylogenetic interest Thus, the vessels in the third visceral arches become the roots of the internal carotids, those of the fourth form a pair of aortae in amphibians and reptiles, or one of this pair an aorta in birds and mammals, whereas those in the fifth pair give rise to the pulmonary arteries The pulmonary arteries are really the sixth pair of aortic vessels because a vestigial fifth pair occurs, together with the sixth, in the fifth visceral arches

In fishes, the aorta in each of the arches which bear gills, commonly four pairs, becomes altered and augmented to form the afferent and efferent vessels of the gills of that arch The precise method of alteration and final condition varies somewhat in different groups of fish

In the elasmobranchs the ventral two thirds of the original aortic arch persists as the afferent vessel This becomes disconnected from the dorsal third from which two vessels grow ventrally These and the dorsal third thus become the efferent aortic arteries and are connected with the afferent vessel through gill capillaries in a manner to be described below (Fig 2*b*) The more anterior of the two growing vessels drains the posterior half of the arch preceding it, while the posterior half drains the anterior half of the arch following In each arch the two efferent vessels have transverse connections midway between them and all the efferent vessels on a side are eventually united by a ventral vessel The teleost condition represents essentially the reverse of what occurs in the elasmobranchs with respect to the origin of the afferent and efferent vessels Thus, in the teleost it is a new upgrowth from the base of the original arch which comprises the afferent vessel whereas most of the original aortic arch forms the efferent vessel In this case also there is only one efferent vessel in each visceral arch (Fig 2*d*) A situation which is intermediate with respect to the origins and character of these vessels will be found in Fig 2*c*, a condition occurring in *Acipenser*, *Amia*, and *Lepidosteus* although in detail these situations are not precisely as shown in the diagram

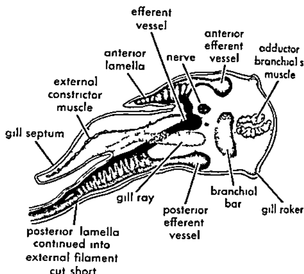


Fig 3 Section across gill bar of *Scyllium canicula* late embryo 32 mm long, showing blood supply to lamellae *(From E S Goodrich, Studies on the Structure and Development of Vertebrates, reprint, 2 vols, Dover, 1958)*

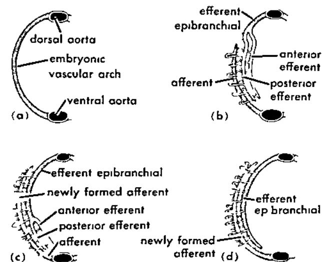


Fig 2 Diagrams illustrating development of adult branchial vessels in various fishes (a) Original continuous embryonic arch cross lined (b) Selachian (c) Intermediate form such as *Acipenser* (d) Teleost, newly developed vessels, white In (b), (c), and (d) the original arch is interrupted *(From E S Goodrich, Studies on the Structure and Development of Vertebrates, reprint, 2 vols, Dover, 1958)*

Branchial gills. The term branchial gills is used because there are gills and structures functioning as gills which are not always in the pharyngeal region There are generally considered to be two kinds of branchial gills, external and internal Internal gills are partly distinguished from external gills by a cover or operculum Because the teleostome internal gill is apparently derived from the elasmobranch type the elasmobranch gill will be discussed first

Elasmobranch gill. The mesodermal core of each branchial arch in these fish soon develops within it a delicate cartilaginous bar slightly proximal to the vessels in addition to the median afferent and two lateral efferent vessels already mentioned The arches also become supplied by branches of the ninth and tenth cranial nerves (the hyoid by the seventh) and acquire muscle fibers derived from the walls of coelomic extensions with which the arches are temporarily invaded From the outer side of each arch there now grows along its dorsoventral extent a thickish sheet of tissue which develops along its border two indentations or clefts, tending to divide the single sheet into an anterior, median, and posterior layer The indentations never become very deep, however, so that these layers are essentially one The middle layer or part of this sheet becomes the septum, consisting of connective

Another peculiar structure used for respiration is the tail fin of an Indian Ocean fish which is immersed while the fish basks out of water One of the Dipnoi, *Lepidosiren*, despite its lungs, apparently develops respiratory filaments on the paired fins. One of the siluroids practices rectal respiration by sucking in and expelling water from the anus.

Gills in Amphibia. Almost all amphibians have gills at some time during development, and some times as adults. At least part of the time these are of the true external type as compared with the filamentous variety previously mentioned Although some authorities consider that the differences between the internal gills and what are termed true external gills are not very significant, it seems that in this class of animals these differences are sufficient to demand some notice Generally speaking true external gills in amphibians may be distinguished from the gills, internal or external, already discussed in that they arise from the outer borders of the gill arches rather than the sides, they differ in certain details of structure, and they are covered by an operculum

Each gill of the type under consideration is composed of a rather heavy main stem or rachis which arises first. This is much thicker than the filaments

In those amphibians in which external gills are

the rachis, at least, are muscle fibers so that the gills can be moved The rachis also contains an arterial loop with extensions into each branch

arches those from the first pair are the most prominent, overlapping and concealing the other two pairs

These external gills in the anurans and many urodeles soon begin to be absorbed and are covered

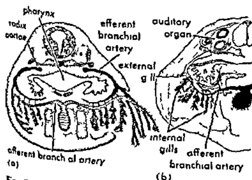


Fig 7 (a) Relations of external and internal gills (b) External gill development (After Maurer from E S Goodrich, Studies on the Structure and Development of Vertebrates reprint, 2 vols, Dover 1958)

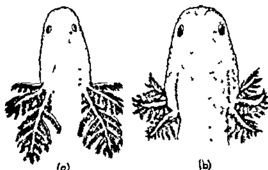


Fig 8 Head and gill form (a) *Pseudobranchus striatus* (b) *Siren lacertina* Drawn from living specimens approximately 6 in in total length (From G K Noble, The Biology of the Amphibia, Dover, 1954)

by back growths of tissue from the hyoid arches, which are continuous across the ventral side of the throat This tissue contains no cartilaginous or bony plates and constitutes an operculum which fuses with the body wall everywhere posterior to the gill slits except for one or two small openings termed spiracles through which water from the clefts leaves the branchial chamber These spiracles are not at all homologous with those previously mentioned

Meanwhile as the operculum is completed the external gills are absorbed, and in the Anura double rows of outgrowths on the borders of the first three pairs of branchial arches and a single row on the anterior border of the fourth are developed in their stead These processes are quite similar to the processes which arose from the main

highly filamentous, and lack any part of a septum Hence they may be regarded as reduced external gills which are covered (Fig 7) There are, nevertheless parts connected with these gills which appear to be closely related to similar parts connected with the internal gills of fish The inner margins of the gill arches bear double rows of papillae which correspond at least in function (acting as strainers), to the gill rakers of the former group

In certain of the urodeles such as *Necturus* and others the original gills are never covered nor absorbed but persist throughout life as actual external structures In these cases, however, they differ somewhat from the larval external gills just described The rachis is larger and heavier and the secondary branches give rise to numerous fine short, tertiary outgrowths giving the whole structure a bushy effect (Fig 8) Although apparently unnecessary some of the clefts between these sets of gills remain open for example one pair in *Pseudobranchus* two pairs in *Necturus* and three in *Siren* This retention of the gills and clefts in such forms has been regarded as an example of neoteny, that is the persistence of larval characters in sexually mature animals See NEOTENY

as in the elasmobranchs, and occur as hemibranchs on the sides of the septa near the pharynx where the tubes become enlarged into spherical pouches The pouches open directly either to the pharynx or through very short ducts The pharyngeal region becomes divided during development in the lampreys into an upper esophageal portion and a ventral respiratory tube with a blunt posterior ending It is this ventral portion which gives rise to the gill slits The number of slits, eventually pouches and ducts, varies greatly in this class ranging from 6 to 14 in different species of *Bdellostoma* (Fig 6)

Teleostome gill. With a few exceptions to be noted below, it is possible to describe the gill situation in most of the teleostomes by indicating the ways in which it differs from that of the elasmobranchs Although the holobranchs start to develop from the four branchial arches, as in the latter group the teleostomes generally lack an open spiracle and the hemibranch related to it Instead there is often a so called pseudobranch which is probably glandular, usually no temporary filamentous external gills occur, the septum grows out only a very short distance, and the anterior and posterior hemibranchs continue to grow out not as sheets but as numerous free filaments corresponding to the primary lamellar folds (Fig 4*b*, *e f*) The secondary lamellae develop in the same way from these and with the same histological character as in elasmobranchs Although there is only a single efferent artery in the arches of most teleostomes (Fig 2*d*) the circulation in the filaments and secondary lamellae is similar to that in the primary and secondary lamellae of the former class Another difference is that instead of a single ray extending between each primary lamella of the posterior hemibranch and the septum there is a ray in each pair of filaments of a holobranch (Fig 4*b*, *e*) Lastly there is a single operculum on each side, attached anteriorly to the hyoid arch and covering all the gills Each operculum is a sheet of tissue in which are embedded three flat bones the operculars Gill rakers are present in one or more (often two) rows, sometimes supported by ossifications (Fig 6*d*)

There are a few exceptions to the situation just described The gills of *Acipenser* are about half way between those of elasmobranchs and typical teleostomes, whereas the gill filaments of the lophobranchs consist of tufted processes

Dipnoian gill. The Dipnoi may be considered as a sort of connecting link between the fishes and the amphibians In Dipnoi there are both internal and apparently true external gills These latter gills exist in the embryos of all the Dipnoi and vestiges of them persist in the adult where they are attached to the last three pairs of arches The internal gills in this group are reduced in correlation with the accessory respiration furnished by the lung or lungs The septa are somewhat diminished causing the primary lamellae to become partly filamentous but without rays The two to four pairs of holobranchs sometimes have hemibranchs on the hyoids on the last pair of branchial arches, or both

Other respiratory devices of fish In addition to pharyngeal gills of the types indicated there are also other aquatic respiratory mechanisms which sometimes occur in quite different locations One such organ although still associated with the pharynx, is found in *Anabas* the climbing perch In this fish some of the pharyngeal bones are developed into folded plates covered with vascular epithelium through which respiration can occur This arrangement is covered by the operculum and can be kept moist for extended periods out of water In one of the physostomes *Amphinous*, another pharyngeal derivative acting as a gill consists of vascular sacs opening out through the spiracles

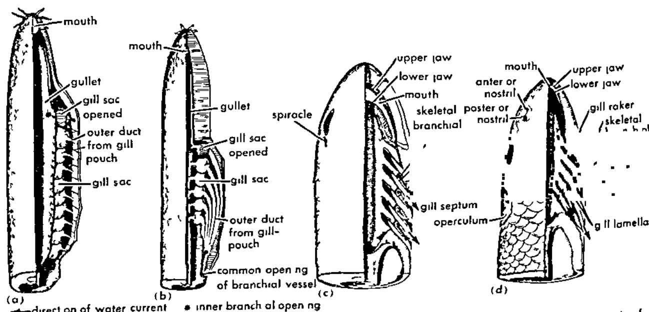


Fig 6 Diagrams of head (*a*) *Bdellostoma* (*b*) *Myxine* (*c*) A selachian (*d*) A teleost Dorsal view partly dissected to show arrangement of gills (*From E S Goodrich, Studies on the Structure and Development of Vertebrates reprint 2 vols, Dover, 1958*)

of the gill surface and in the dimensions and the spacing of the lamellae. Such differences are largely associated with adaptive radiation Active pelagic species tend to have large, finely divided surfaces while more sluggish demersal species tend to have less lamellar surface and somewhat coarser channels. Those fish which have organs for breathing air in addition to gills have reduced gill surfaces *There are also wide quantitative differences in* the oxygen capacity of the blood and in the nature of the loading curve of the hemoglobin which also appear to be of adaptive significance Active fish *tend to have* blood of higher oxygen capacity, *the* blood of those which live in well aerated waters requires a relatively high partial pressure of oxygen for saturation. There are also major differences in the sensitivities of fish bloods with respect to the effect of the presence of carbon dioxide on the oxygen capacity These, too, probably have their part in fitting the various species for the environments in which they are found

The major phylogenetic change in the gill system has been in the means of irrigation In the marsipobranchs and the elasmobranchs there is a series of individual branchial chambers to the walls of which the gill filaments are fixed for the major length of their postbranchial margin Each chamber also has its own external aperture with its own valve (except in the chimaerids) Thus, in these groups there is a very restricted postbranchial chamber In the higher fishes the gills are no longer in a series of separate chambers, the walls have been reduced to an arch which bears the blood vessels and the supporting cartilages, and a new structure, the operculum and the branchiostegial apparatus, has [illegible]

[illegible] largely by pressure exerted by *decreasing the* volume of the oral cavity In the higher forms with the extensive postbranchial chamber, water is drawn through the gills by the expansion of this chamber as well as forced through them by the constriction of the oral cavity These two processes of suction and pressure are synchronized to permit the intake and expulsion of the respiratory water to take up but a minor fraction of the total respiratory cycle and thus to allow an almost continuous passage of water over the gills The freeing of the filaments from the walls also makes for a less restricted passage for the respired water The suction pump appears to be most highly developed in demersal species When fish are swimming, water *may be passed over the gills simply* by the fish's leaving the mouth and the external branchial passages open Thus, in pelagic species, the pumping apparatus, whatever its nature, may be reduced It is possible that any advantage the more efficient pumping system may confer on the higher fishes operates only under conditions of rest or when they are recovering from exercise [F E J F]

EMBRYOLOGY OF THE LUNG

Lung. The lung is an organ adapted to respiratory exchange of gases between air and the blood of vertebrates It ranges from the simple swim bladder of dipnoid fish to the large spongy, compound air sac of man The story of the lung includes the evolution of certain vital accessories to respiration, the diaphragm and thoracic wall All lungs correspond in originating as a pocket from the pharynx near the level of the sixth aortic arches which furnish the blood supply, in innervation by the vagus nerve and in their embryonic position in the abdomen This position is retained in the adult of fish amphibians, and reptiles and partially retained in birds but *in mammals the lung is segregated* from the abdomen in fetal life into a new formed thoracic cavity (Fig 9*a*, *b*, and *c*) *See* PHARYNX, SWIM BLADDER

Extrusion of the lung. *In the* extrusion of the lung from the abdominal cavity it first invaginates *the cervical aponeurosis of the abdomen* and in so doing enters the space being formed between the *inner and middle layers* of the body wall, the pleural

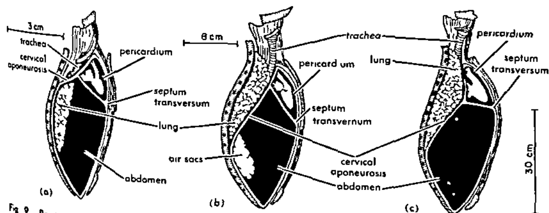


Fig 9 Position of the lung (a) Fish and amphibians (b) Birds. (c) Mammals (After Keith by permission of J D Boyd and the Cambridge University Press)

Other amphibian respiratory devices. In the amphibians as in fish there are various peculiar structures which act as gills Thus, in the marsupial frog the two anterior pairs of gills are transformed into vascular wrappings which surround the body, whereas the balancers of some salamanders have been said to have a partly respiratory effect, al though this is doubtful The skin in amphibians certainly has a respiratory function, and that of *Cryptobranchus* is thrown into vascularized folds which are waved about Also in the hairy frog the skin of the breeding male develops filamentous processes which aid in respiration

Gill ectoderm and endoderm. It has been claimed that the epithelium of internal gills is endodermal, but because the boundary between endoderm and ectoderm is indistinguishable by the time the gills arise, this is difficult to prove It has also been stated that the ectoderm grows inward and covers the areas from which the gill lamellae and rakers develop, thus making the epithelium of these parts ectodermal This would at least account for occasional rakers with enamel a substance sup posed to be derived only from ectoderm The point of origin of gills designated as truly external marks their epithelium as clearly ectodermal, this is also probably true of the later internal gills of those amphibians which have them [R S M]

COMPARATIVE PHYSIOLOGY OF GILLS

The functions of gills in fish are considered be cause these animals have been studied more ex tensively than others The gills of fish serve for the exchange of materials between the blood and water They are permeable to certain substances of im portance to the economy of the organism, notably oxygen, carbon dioxide, and the nitrogenous wastes ammonia and urea These are transferred through the gill by simple diffusion In addition the gills are somewhat permeable to water a circumstance for which most fish must compensate by the active transport of certain ions to maintain their osmotic balance A significant part of this latter function is also carried out by the gills The gills cannot be considered by themselves alone, because their func tioning depends as much on the systems which bathe their external surface with water and their internal surface with blood as it does on the gills themselves

Gill filament. The fundamental unit of the gill is the gill filament The filaments are long flattened slightly crescent shaped structures extending later ally from openings in the side of the pharynx Fila ments on the opposite sides of a given pharyngeal opening arch towards each other, at least at their lateral extremities, their tips meeting and thus enclosing a space between the filaments and the pharyngeal opening that may be termed the pre branchial chamber External to the arched free portions of the filaments is a postbranchial chamber that communicates with the exterior through an opening in the body wall

Lamellae The gill filaments bear two important types of structure The first of these is the second ary lamella which provides the actual surface for diffusion The lamellae are almost microscopic, somewhat semicircular, leaflike extensions trans verse on the flat faces of the filament They consist of a loose web of supporting tissue so that there are extensive blood lacunae within them They are covered by an epithelium of a single layer of cells, a circumstance which provides the least barrier to diffusion The lamellae do not extend completely across the filament but stop short of its post branchial edge Thus, even in those groups in which the filament is fixed, a septum extends along most of its postbranchial margin and there is a channel beyond the lamellae that leads to the post branchial chamber There are 10–30 secondary lamellae per millimeter of gill filament, and the lamellae of adjacent sides of two neighboring fila ments interdigitate The effect of the whole array of filaments is that of a mesh of tiny channels between the interdigitated lamellae Water is pumped through this mesh from the prebranchial to the postbranchial chamber

Blood enters the lamella at the postbranchial end of its base and leaves at the prebranchial side where it is collected by the efferent blood vessel, a prominent feature in gills as ordinarily seen in injected specimens in the anatomy laboratory Blood and water thus travel in opposite directions as the blood flows through the lamellae and the water through the tiny spaces between them This arrangement provides for a highly efficient transfer system Up to at least 80% of the oxygen content of the water may be removed as it passes over the gills of a quietly breathing fish It may be pre sumed that the other diffusible substances are transferred with similar efficiency

Chloride excreting cell The other type of struc ture of importance to the special functions of the gill is the so-called chloride-excreting cell These are columnar acidophilic cells that are concen trated in the postbranchial region of the filament and associated with the afferent blood supply They are clustered about the bases of the lamellae and occur on the face of the filament itself

Osmotic and ionic regulation. There is a con stant loss of water in marine fish except for the hagfishes and the elasmobranchs in which, for different reasons the blood is isosmotic with sea water The loss is made up by swallowing sea water Univalent ions absorbed in the sea water are excreted through the gills, and the agent for this excretion is considered to be the chloride excreting cell In fresh water the osmotic shift of water is in the reverse direction In this case, the excreting cells reverse their polarity and transfer chloride from water to blood It has been suggested that there may actually be two types of chloride cell at least in the cyclostomes, one type regressing and the other type developing as the fish passes from salt to fresh water, rather than the reversal of the action within the same cell

Branchial system. There are comparative dif ferences in each component of the branchial system There are wide quantitative differences in the area

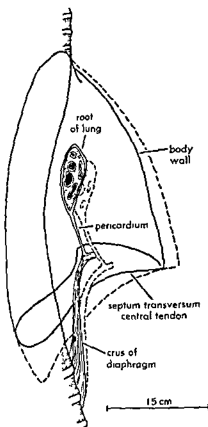


Fig 13 Diagram showing the actual respiratory movements of the diaphragm, pericardium, and root of the lung in higher mammals. Broken lines show change in position of thorax and viscera during inspiration (*After Keith by permission of J D Boyd and the Cambridge University Press*)

Amphibian lung. The amphibian lung is bifid and each half projects into the abdomen above the pericardium and liver (Fig 10) The inner surface of the lung present ...

... (Fig 15), extends forward to the cervical aponeurosis. The 8th to 14th ribs are complete, and three cervical ribs can both expand and compress the apex of the lung The sternum is a strong fulcrum for the costal cartilages The cervical part of the trunk is increased in length so that the head and pharynx are carried away from the heart Whereas in amphibians the swallowing of air tends to force ...

... into the lungs

... through the cervical apone...

... these into

two layers, the dorsal or pulmonary, and ventral or abdominal (Fig 16) If these lungs were removed from the thorax and the avian septum were replaced against the dorsal wall, the 3 layered body wall of the amphibian would be restored Although the lungs are in the thorax, the air sacs remain in the abdomen.

Human embryo. The lung in the 5 week human embryo lies in the abdomen The septum transversum in man consists of a ventral part (Fig. 17*b*,*c*) representing the entire septum of amphibians (Fig 11*a*), and a dorsal part representing a fusion of the amphibian cervical aponeurosis with the dorsal wall of the pericardium The common cardinal vein (Fig 17*b*,*c*) runs in the anterior part of the septum next to the wall of the pericardium The lung (Fig 17*b*,*c*) lies medial to the Wolffian fold which is attached to the mesonephros dorsally, the liver ventrally, and the septum in front with its free border looking backward In the amphibians the lungs grow backwards in the abdomen and are attached to the mesentery between the aorta and esophagus above and liver and stomach below. In the human embryo they advance medial to the septum (Fig 17*b*,*c*), then evaginate laterally into the dorsal part of the septum between its anterior part which contains the common cardinal vein next to the pericardial roof, and the posterior part which develops to become the pleuroperitoneal membrane

After this extrusion of the human lung from the abdominal cavity, the lung lies next to the medial surface of the middle layer of the body wall Upon removal of the body wall, the lungs are exposed

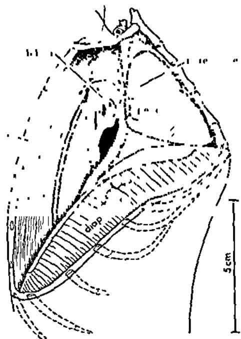


Fig 14 Diagram showing discontinuity of the diaphragm with the pericardium and hilum of the lung in the typical mammal (rabbit) (*After Keith by permission of J D Boyd and the Cambridge University Press*)

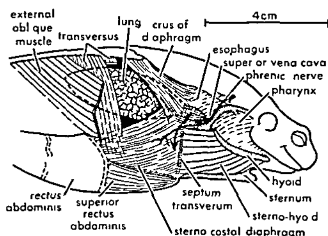


Fig 10 Diagram showing lung in an amphibian (*Surinam toad*) and muscles from which the diaphragm is evolved (*After Sir Arthur Keith by permission of Edward Arnold Ltd*)

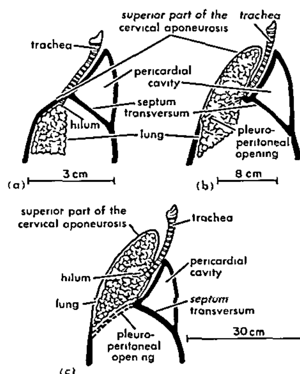


Fig 11 Lung extrusion in (a) amphibians, (b) birds, (c) mammals (*After Keith by permission of J D Boyd and the Cambridge University Press*)

cavity These events bring the lung to an anterior thoracic position, and also allow it to develop an apex, both important in respiration In amphibia (Fig 11*a*) the lung comes to lie in the pleuro peritoneal cavity It is infrapericardial, with both the root and apex in the same plane Figure 11*b* represents the condition in birds in which an apex grows *cranially beyond the root of the lung* Thus, the entire lung remains abdominal, because the apex carries forward the entire cervical aponeurosis The human lung becomes supradiaphragmatic (Fig 11c). The diaphragm is represented by only a thin layer of the aponeurosis that is carried for ward (Gibson's fascia) and the remainder of the aponeurosis serves to close the passageway Moreover, the root also is carried forward, and in man it becomes closely bound to the pericardium (Fig 12)

The pericardial bond is of primary importance in man and the great anthropoids because the pericardium is bound also to the central tendon and diaphragm (Fig 12) The part of the lung above the pulmonary root is relatively large, and the apex, located at a fixed point at the level of the neck of the first rib, can expand when the diaphragm contracts and pulls the root and heart in a downward and forward direction (Fig 13) Upon expiration they move in the reverse direction In the typical mammal, there is little or no contact of the diaphragm with the pericardium, and the part of the lung above the root is relatively small (Fig 14) *Here the descent of the diaphragm has little effect* on the movement of the heart and pulmonary roots. Instead, the azygos lobe expands between the esophagus and inferior vena cava into the subpericardial space beside the pericardial mesentery

The body wall becomes an inspiratory mechanism in reptiles, as a result of the ribs sternum, and intercostal muscles which evolve in the middle layer of the wall In phylogeny the first intermuscular septum to extend completely to the median ventral line is that behind the seventh segment of the rectus abdominis in the amphibian *Necturus maculatus* In this septum the first sternal rib develops in typical reptiles and mammals

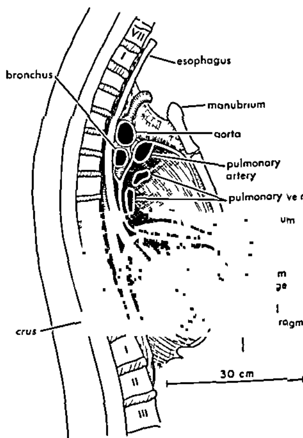


Fig 12 Diagram to show connections of the diaphragm, pericardium, and root of the lung in man (*After Keith by permission of J D Boyd and the Cambridge University Press*)

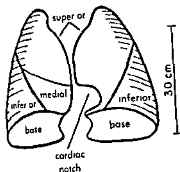


Fig 18 The lungs anterior aspect (After Grant by permission of the Williams and Wilkins Company)

cava in front of the pulmonary ligament Other impressions shown in the figures are those of the esophagus and trachea, the left subclavian artery, and brachiocephalic or innominate vein The borders of the lung are pinched extensions between the pericardium and body of the sternum (anterior border) and between the diaphragm and body wall (inferior border) The inferior limit of the thoracic cavity on both sides is related to ribs 8, 10, and 12, whereas the lower borders of the lungs stop at ribs 6 8 and 10, and the unoccupied space is the costodiaphragmatic recess on both sides The cardiac notch is the absence of lung because of pressure from the heart A similar but smaller, notch is made in the underlying pleura, which gives rise to the left costomediastinal recess

The oblique fissure cuts through the costal, diaphragmatic and medial surfaces to the root of the lung (Fig 18) In the right lung the horizontal fissure runs backward from the anterior border and meets the oblique fissure in the midlateral line Thus, the right lung has three lobes, superior middle, and inferior, whereas the left lung has two, the superior and inferior The lingula, or antero inferior part of the left upper lobe and the cardiac notch above it correspond to the middle lobe of the right lung (Fig 18)

Bronchopulmonary segments. For convenience in exploration and study of the lung, it may be divided into anatomical areas The bronchial tree branches mainly by dichotomy (Fig 20) The ultimate generations, that is, the respiratory bronchioles, alveolar ducts, and alveoli constitute all of the respiratory portion of the lung This respiratory portion (Fig 21) consists of 10 segments in the right lung and 8 in the left, each of which is supplied by a tertiary branch of the branchial tree (Fig 20) Two or more of the segments make up a lung lobe Further, the primary bronchi (from the trachea) divide as secondary bronchi, three on the right side and two on the left, corresponding to the lobes of the lung Because the upper left lobe results from the fusion of two lobes, the prospective

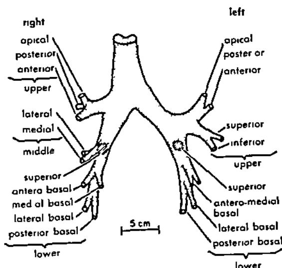


Fig 20 The 10 right and 8 left segmental bronchi (After Jackson and Huber by permission of the Williams and Wilkins Company)

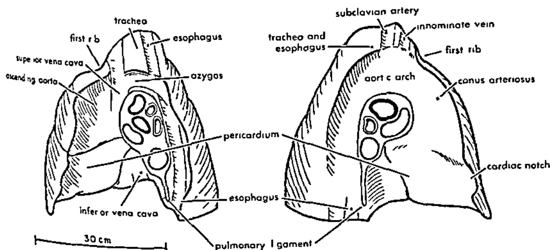


Fig. 19 Impress ons on the mediastinal surfaces of the lungs. After Grant by permission of the Williams and Williams Company)

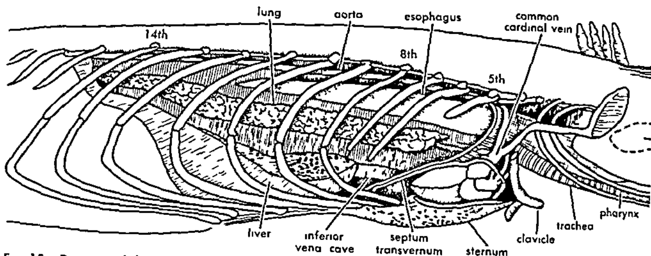


Fig 15 Diagram of the reptile, *Iguanodon* sp, showing the abdominal lung (*After Keith by permission of J D Boyd and the Cambridge University Press*)

The lungs can grow in a forward or backward direction, and ventrally between the inner and middle layers of the body wall pari passu as these open up The site of extrusion (pleuroperitoneal opening) need only close, and the lung will occupy new, completely enclosed, pleural cavities

ANATOMY

The shape and volume of the lung, because of its pliability, conforms almost completely to that of its cavity The lungs are conical (Fig 18), each has an apex and a base two surfaces, two borders, and a hilum

Thoracic cavity. The apex extends into the superior limit of the thoracic cavity The base is the diaphragmatic surface (Fig 18) The costal surface may show bulgings into the intercostal spaces The medial surface has a part lying in the space beside the vertebral column and a part imprinted by the form of structures bulging outward beneath the mediastinal pleura (Fig 19) The hilum and pulmonary ligament descending from it are notable The cardiac impression is deeper on the left lung because of the position of the heart The aorta arches over the left hilum and the azygos vein over the right Joining the right cardiac impression are the groove for the superior vena cava in front of the hilum and that for the inferior vena

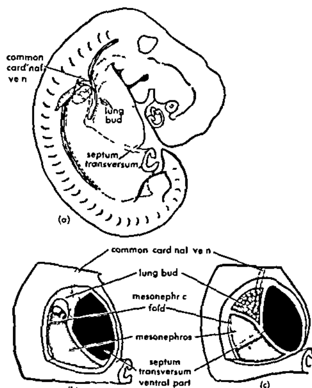


Fig 17 (a b) Diagram of human embryo of 5 weeks to show initial process of lung extrusion (c) Human embryo of about 7 weeks showing extrusion into the cervical aponeurosis (*After Keith by permission of J D Boyd and the Cambridge University Press*)

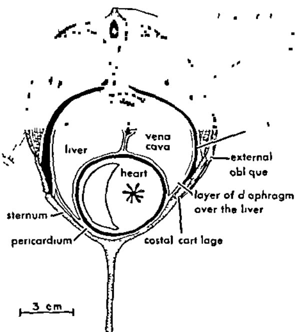


Fig 16 Diagrammatic cross section of the thorax (lung), abdomen, and pericardium in the pigeon (*After Keith by permission of J D Boyd and the Cambridge University Press*)

[...]ers. Efferent sympathetic fibers are bronchodila[...]tor hence the use of adrenalin for relief of b[r]on[c]hial [s]pasm resulting from asthma

HISTOLOGY

E[ss]ential to lung function histologically are the respiratory passages pleura blood vessels, lym[...]phatics, and re[s]piratory surfaces proper

Epithelium Between the cartilage in the walls and the surface epithelium of the air passages are

[...]ncrea[s]e relatively in amount with decrease in size of the bronchioles In the terminal bronchioles wh[i]ch are free of cartilage smooth muscle is the main const[i]tuent and may close the bronchiole as in a[st]hma. The epithelium is the respiratory type that is the cells are tall and capped by motile c[i]l[i]a. The cilia wave toward the throat carrying along the secretions of the glands laden with in[s]pired du[s]t

Pleura. The pleura lines the outer surface of the lung and the inner surface of the pleural cavity It consi[st]s of a layer of flattened mesothelial cells and an underlying layer of collagenous and elastic fibers which support numerous blood and lymphatic cap[i]llaries, fibroblasts and macrophages The pleura cont[i]nuously pours out a mucoid exudate which lubricates the opposed surfaces The pleura [i]s pink [i]n the newborn child

Respiratory portion This portion includes the respiratory bronchioles which branch from the term[i]nal bronchioles the respiratory ducts alveo[...]lar sacs and alveoli. The respiratory ducts are sim[i]lar to terminal bronchioles except that they are

[...]ous alveoli so that the openings of the alveoli oc[...]cupy the greater area of the wall Alveolar sacs are term[i]nal dilations of the alveolar ducts Alveoli are the smallest functional units of the lung Their walls are made up of one or two capillaries a few ela[s]tic fibers and more reticular fibers among which are a few fibroblasts dust cells or lung mac[r]ophages and septal cells The septal cells give rise to some of the macrophages and others are carried [i]n by the blood Macrophages from both sources wander [...]

[...]e [d]emonstrated with ordinary techniques The grea[t]er volume of the lung is taken up by air space of the respiratory portion Thus the lung feels spongy and in sections looks like fine lace [L P C]

PHYSIOLOGY OF THE LUNG

Simple molecular d[i]ffusion of gases underlies respiratory exchange in all animals However the lungs of air breathing vertebrates require mechani[c]al ventilation to sustain this diffusion The molec[...]ular characteristics of respiratory gases (Table 1) the physical properties of atmosphere (Table 2) and of body fluids (Table 3) the physiological lim[...]itations on lung structure and location and the large requir[...] (O_2) uptak[e] (Table 5) [...] respiration alone Diffusion suffices only within a few millimeters of those membranes across which interchange of O_2 and CO_2 can occur Refinements in ventilation respiration intervene as continuing speciation involves increased O_2 requirements The functional as well as structural changes follow rec[...]ognizable evolutionary trends See METABOLISM, RESPIRATION

Other vital requirements Besides exchange of O_2 and CO_2 there are other vital requirements for breathing processes in air Among these are (1) limitation of the evaporation of water in general and control of the dissipation of heat in birds and mammals in particular (2) ventilation of the ol[...]factory membrane (3) production of high intra[...]pulmonary pressure for such acts as coughing def[...]ecation and parturition and low intrapulmonary pressure for regurgitation (4) provision for large pulmonary volume in such acts as sneezing yawn[...]ing and phonation and (5) allowance for buoy[...]ancy control in aquatic animals All these ancillary processes involve corresponding variations in the respiratory mechanisms As animals are compared differences are found (1) in buccopharyngeal pulmonary interstitial and diaphragmatic muscula[...]ture for inducing and directing air movement (2) in valvular muscle of nares glottis and pul[...]monary ducts or apertures for controlling flow and pressure (3) in accessory air sacs to increase ca[...]pacity for ventilation itself or for phonation (4) in heart for adequate pulmonary blood supply and in circulation for efficient perfusion (5) in brain and peripheral nerve patterns for initiation main[...]tenance and adjustment of ventilatory processes control of circulation and regulation of ancillary functions and (6) in skeletal and connecting structures A brief review of comparative respira[...]tion can only delineate the basic trends in phylog[...]eny of respiration Certain general features of the breathing process must first be considered before comparisons among species can be made See SPEECH

Breathing The act of breathing induces mass flow of air in and out of the respiratory organs This inspiration and expiration mechanically ven[...]tilates the entire lung of those animals such as lungfish salamander and frog in which the organ is a simple sac Here the inner lining or pulmonary epithelium where diffusion itself occurs is directly exposed to the tidal air flow caused by breathing In these animals breathing is intermittent or peri[...]odic that is cycles of inspiration and expiration occur in groups after many minutes of nonbreath[...]ing or apnea The periods of apnea may exceed those of ventilating The lung is normally closed from the atmosphere by a valve (glottal valve) at

left upper and middle lobar bronchi become par tially fused As a result, the four segmental bronchi of the upper lobe are not tertiary, but bronchi of the fourth division of the tree However, there is great similarity in the structure of the right and left bronchi

The trachea and extrapulmonary bronchi are kept open by C shaped bars of hyaline cartilage Within the lung the bars are arcs staggered in the bronchial walls at different levels When in their branching the bronchi and bronchioles are reduced to a diameter of 1 mm or less, they are then free of cartilage, and are called terminal bronchioles One of the terminal bronchioles enters the apex of a secondary lobule of the lung These secondary lobules are anatomic units of the lung whose hexagonal bases 10–20 mm in diameter, rest on the pleura (Fig 21) or next to a bronchiole or blood vessel and whose apices point toward the hilum Finer lines divide the bases of the secondary lobules into smaller areas (Fig 20) These are the bases of primary lobules each served by a respira tory bronchiole

Blood supply. The blood supply to the lung is provided by the pulmonary and the bronchial arteries The right pulmonary artery runs dorsally between the right upper and middle lobe bronchi so that the upper bronchus arises 1 in from the trachea The left pulmonary artery and the arch of the aorta pass dorsally between the trachea and the left upper lobe bronchus so that the latter is 2 in from the trachea Each pulmonary artery divides into 10 branches which follow closely the postero superior walls of the segmental bronchi They take the names of the 10 right segmental bronchi

The bronchial arteries arise on the left side from the aorta and on the right from either an intercostal

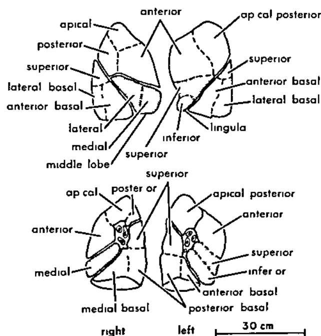


Fig 21 The 10 right and 8 left bronchopulmonary segments (After Jackson and Huber by permission of the Williams and Wilkins Company)

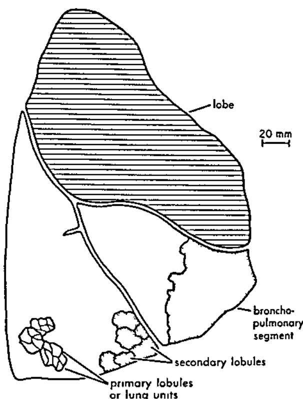


Fig 22 The subdivisions of the lung (After Grant by permission of the Williams and Wilkins Company)

or the left bronchial artery They supply the walls of the bronchi pulmonary vessels and lymph nodes They pass with the radicles of the pulmonary vein through the interlobular septa and supply the pulmonary pleura Blood delivered to the lung by the bronchial arteries is returned by the radicles of the pulmonary veins except that to the largest bronchi which is returned by the bronchial veins to the azygos veins

The pulmonary veins have 10 branches, their main stems run in the medial or inferior sides of the bronchi Because their tributaries run inter segmentally and drain adjacent segments, and because the arteries may cross intersegmental bound aries a bronchopulmonary segment would not be a morphologic bronchovascular unit The upper and lower lobes of the left lung send an upper and a lower pulmonary vein respectively, into the left side of the left atrium The upper and middle right lobes provide the upper vein to the right side of the atrium

Lymphatic channels are not found in interalveolar partitions but occur everywhere else in the lung except in cartilage There are two main sets of lung lymphatics, those in the pleura, and those within the lung The latter begin at the alveolar ducts and follow the bronchi and pulmonary vessels to the lymph nodes at the hilum Those in the pleura form secondary and primary networks in the lobular septa and drain into the hilar lymph nodes

Nerve supply. The nerves which supply the lung are branches of the vagus, and of the thoracic sympathetic ganglia 2, 3, and 4 Efferent vagal fibers are bronchoconstrictor and secretory, whereas the afferents are part of the arc for the breathing re

Table 3. Respiratory exchange characteristics in man[a]

Part I Ventilation

Gas	Inspired air		Alveolar air[b]		Expired air	
	Composition, vol %[c]	Partial pressure, mm Hg[d]	Composition, vol %	Partial pressure, mm Hg[e]	Composition vol %[c]	Partial pressure, mm Hg[e]
H_2O	0.00	5.7	00.0	47	00.0	47
N_2	79.02	596.0	80.4	573	79.2	565
O_2	20.95	158.0	14.0	100	16.3	116
CO_2	0.03	0.3	5.6	40	4.5	32

Part II Transport[f]

Gas	Arterial		Capillary		Tissue fluid		Venous	
	vol %	mm Hg	vol %	mm Hg	vol %	mm Hg	vol %	mm Hg
H_2O	83 (81–86)	47	83 (81–86)	47	83 (81–86)	47	83 (81–86)	47
N_2	0.975	573	0.975	573	0.975	573	0.975	573
O_2	19.6 (17.3–22.3)	94	1–22.3[g]	1–94[g]	0.185[g]	30[g]	12.9 (11.0–16.1)[h]	40
CO_2	48.2 (44.6–50.4)	40	44.6–57.7[g]	40–50[g]	3.016[g]	50[g]	54.8 (51.0–57.7)[h]	46

example) involves trunk musculature to supply pulmonary ventilation in all reptiles birds and mammals. A rather continuous buccopharyngeal ventilation not always involving pulmonary ventilation itself is characteristic of amphibians (Table 4) and persists in lizards, turtles, and other reptiles In amphibians, therefore buccopharyngeal activity subserves both breathing and olfaction, in reptiles it subserves only the olfactory sense (smell)

Table 4 Some respiratory values in vertebrates*

Animal†	Weight kg	Breathing rate cycles/min	Tidal volume ml	Minimum volume liter/min	O_2 consumption $mm^3/(g)(hr)$	Intrapleural pressure, cm H_2O	Compliance liter/cm H_2O
Frog (*Rana fusca*)					210.0 (20°C)		0.001
Turtle (*Malaclemys centrata*)		3.7 (24°C)	14.0	0.051	35.0 (24°C)		
Alligator (*Alligator mississippiensis*)					8.9 (22°C)		
Canary (*Serinus canarius*)		108			2900.0		
Chicken (*Gallus domesticus*)		17	45.0		497.0		
Duck (*Anas sp.*)		42	36.5		800.0		
Rat (*Rattus norvegicus*)	0.273	60	1.4	0.074	770.0	−2 to −8	0.0012
Dog (*Canis familiaris*)	20	17	302.0	5.30	580.0	−5.4 to −13.5	0.09
Horse (*Equus caballus*)	696	12	9060.0	107.0	250.0	−8.0 to −22.0	0.80
Man (*Homo sapiens*)	66	14	372.0	5.04	220.0	−3.8 to −9.3	0.20

* From D. S. Dittmer et al (eds.) *Handbook of Respiration* 1958 and H. H. Dukes, *Physiology of Domestic Animals* 1955

† Adult male at rest.

Table 1 Characteristics of the respiratory media*

Variable	Water		Atmosphere (N_2)	
	Ocean	Fresh	Sea level	6000-m altitude
Temperature, °C	−2 0 to 30 0	2 0–32 0	0 7–15 7	−28 1 to −15 1
Pressure, total mm Hg	760–760,000	760–20,000	760	347 5 to 360 2
Density, g/liter	1027[b] (20°C)	1000[b] (4°C)	1 223–1 290	0 649–0 659
pH	7 5–8 4	3 2–10 6		
Concentration, vol %				
H_2O	100 00	100 00	1 00[c]	1 00[c]
N_2	1 03[b] (15°C)	1 33[b] (15°C)	78 03 (STP)	78 03 (STP)
CO_2	0 02[b] (15°C)	0 03[b] (15°C)	0 03 (STP)	0 03 (STP)
O_2	0 58[b] (15°C)	0 72[b] (15°C)	20 99 (STP)	20 99 (STP)
Salts	3 46[b]	0 18[b]		
Inert gases	Trace	Trace	0 95 (STP)	0 95 (STP)
Partial pressure (tension), mm Hg				
H_2O	12 79 (15°C)	6 10 (4°C)	6 40[d] (15°C)	0 72[d] (−15°C)
N_2	593 02 (STP)	593 02 (STP)	593 02 (STP)	281 06 (STP)
CO_2	0 23[b] (STP)	0 23[b] (STP)	0 23 (STP)	0 11 (STP)
O_2	159 52[b] (STP)	159 52[b] (STP)	159 52 (STP)	75 61 (STP)
Inert gases	7 46 (STP)	7 46 (STP)	7 46 (STP)	3 42 (STP)
Total pressure	760 00	760 00	760 00	360 20
Diffusion coefficient ml/(min)(cm²)(cm) at 760 mm Hg 20°C				
H_2O				
N_2		0 000018 (0 53)[e]		
CO_2		0 000785 (23 1)[e]		
O_2		0 000034 (1 0)[e]	11 0	

[a] From W S Spector (ed), *Handbook of Biological Data*, 1956 [b] Average of many determinations, varies with conditions of measurement [c] Varies, but never absent and always of biological significance [d] Calculated for 50% relative humidity [e] Values in parentheses are relative coefficients with O_2 as unity

the opening into the trachea supplemented by valves (nasal valves) at the nares The lung, however, does not remain a simple sac Beginning with toads and climaxing in birds and mammals more elaborate secondary and tertiary sacculations, the alveoli and the alveolar or air sacs, evolve This evolution increases the diffusion surfaces tremendously and also removes these surfaces from direct exposure to ventilatory air flow Thus, a uniform gaseous exchange medium comes to prevail at the respiratory membranes of birds and mammals the breathing cycles become continuous and shallow (eupneic), and the lung remains normally open to the atmosphere Nasal valves are usually absent in animals above reptiles, and the glottis closes only for such occasional acts as swallowing and coughing *See* RESPIRATION, EXTERNAL

Breathing in vertebrates. Breathing induces air movement resulting from a mechanically imposed pressure difference between a compressible cavity and the atmosphere In mammals and reptiles this cavity is the lung itself In birds and amphibians, however secondary cavities are responsible, namely, special air sacs in the former and the mouth (buccopharyngeal cavity) in the latter These secondary cavities communicate with both the lung and the atmosphere and accommodate the bellowslike action Two basic processes of air breathing reflect the structural differences The more primitive type buccopharyngeal breathing, is found in lungfish and amphibians It involves the same basic neuromuscular elements of mouth and throat as water breathing in fish The more specialized type thoracoabdominal breathing (in man for

Table 2 Characteristics of respiratory molecules*

Type	Weight (O = 16)	Diameter † $cm \times 10^{-8}$	Density g/liter	Mean free path $cm \times 10^{-6}$ (750 mm Hg)	Collision frequency (20°C)	Average velocity cm/sec	Water solubility		Vol % (10°C)
							STP	20°C	
N_2	28 02	3 15–3 53	1 251	8 50	5070	45 100	2 35	1 54	1 18
H_2O	18 02	3 0–5 0	0 0005–0 030‡			56 600			
CO_2	44 01	3 31–3 40	1 977	5 56	6120	36 200	171 3	87 8	53 0
O_2	32 00	2 92 2 98	1 429	9 05	4430	12 500	4 89	3 10	2 31

* From W S Spector (ed) *Handbook of Biological Data* 1956 Unless otherwise indicated values are for standard conditions (STP) of temperature (0°C) and pressure (760 mm Hg) † Range indicates variability with method of measurement (such as viscosity heat conductivity) ‡ Water vapor in saturated air that is in equilibrium with water at 0°C and 30°C

Table 3 Respiratory exchange characteristics in man[a]

Part I Ventilation

Gas	Inspired air		Alveolar air[b]		Expired air	
	Composition, vol %[c]	Partial pressure, mm Hg[d]	Composition, vol %	Partial pressure, mm Hg[e]	Composition, vol %[c]	Partial pressure, mm Hg[e]
H_2O	0 00	5 7	00 0	47	00 0	47
N_2	79 02	596 0	80 4	573	79 2	565
O_2	20 95	158 0	14 0	100	16 3	116
CO_2	0 03	0 3	5 6	40	4 5	32

Part II Transport[f]

Gas	Arterial		Capillary		Tissue fluid		Venous	
	vol %	mm Hg	vol %	mm Hg	vol %	mm Hg	vol %	mm Hg
H_2O	83 (81 86)	47	83 (81–86)	47	83 (81–86)	47	83 (81–86)	47
N_2	0 975	573	0 975	573	0 975	573	0 975	573
O_2	19 6 (17 3–22 3)	94	1–22 3[g]	1–94[g]	0 185[g]	30[g]	12 9 (11 0–16 1)[h]	40
CO_2	48 2 (44 6–50 4)	40	44 6–57 7[g]	40–50[g]	3 046[g]	50[g]	54 8 (51 0–57 7)[h]	46

[b] Alveolar air actually last part of expired

depending on blood flow, tissue activity, and relation of sample to capillary length or field [h] Internal jugular

example) involves trunk musculature to supply pulmonary ventilation in all reptiles, birds, and mammals A rather continuous buccopharyngeal ventilation not always involving pulmonary ventilation itself is characteristic of amphibians (Table 4) and persists in lizards, turtles and other reptiles In amphibians, therefore, buccopharyngeal activity subserves both breathing and olfaction, in reptiles it subserves only the olfactory sense (smell)

Table 4 Some respiratory values in vertebrates[a]

Animal[†]	Weight kg	Breathing rate cycles/min	Tidal volume ml	Minimum volume liter/min	O_2 consumption mm³/(g)(hr)	Intrapleural pressure, cm H_2O	Compliance, liter/cm H_2O
Frog (*Rana fusca*)					210 0 (20°C)		0 001
Turtle (*Malaclemys centrata*)		3 7 (24°C)	14 0	0 051	35 0 (24°C)		
Alligator (*Alligator mississippiensis*)					8 9 (22°C)		
Canary (*Serinus canarius*)		108			2900 0		
Chicken (*Gallus domesticus*)		17	45 0		497 0		
Duck (*Anas* sp)		42	36 5		800 0		
Rat (*Rattus norvegicus*)	0 273	60	1 4	0 074	770 0	−2 to −8	0 0012
Dog (*Canis familiaris*)	20	17	302 0	5 30	580 0	−5 4 to −13 5	0 09
Horse (*Equus caballus*)	696	12	9060 0	107 0	250 0	−8 0 to −22 0	0 80
Man (*Homo sapiens*)	66	14	372 0	5 04	220 0	−3 8 to −9 3	0 20

[a] From D S. Dittmer et al (eds.) *Handbook of Respiration* 1958 and H H Dukes *Physiology of Domestic Animals* 1955

[†] Adult male at rest

Buccopharyngeal breathing Buccopharyngeal breathing is indirect when compared with thoraco abdominal It involves two distinct stages ventila tion of the mouth and ventilation of the lungs The necessary pressure gradients between mouth and atmosphere and between mouth and lungs are gen erated by muscles which raise and lower the hyoid apparatus and floor of mouth and throat the same mechanism is used for water breathing in fish Pul monary inspiration in buccopharyngeal breathing is more descriptively an injection or adspiration The volume of air inspired or expired per breathing cycle is called tidal volume

In the frog movement of air between mouth and atmosphere requires only about 3–5 mm of water (H_2O) pressure that between mouth and lungs reaches 25–35 mm H_2O pressure during the peak of pulmonary inspiration A gradient is directed to-ward the lung of about 20 mm on inspiration be cause a volume of gas (functional residual volume) remains in the lung from the preceding expiration under about 10 mm H_2O pressure The glottis closes at the end of inspiration and a positive intrapulmonary pressure persists during apnea of about 20 mm H_2O as a result of elastic recoil and muscular tonus in lung itself and in body wall These forces also cause expiration when the glottis opens Whether buccal ventilation alone or pul monary ventilation or a combination of these oc curs depends upon neuroregulatory processes which determine the relationship of nasal and glottal valves with each other and with the breathing mus culature Thus a frog s lungs and body can be dis tended greatly beyond normal dimensions by suc cessive inspirations alone

Vital capacity The excess capacity of any ani mal to inspire beyond normal tidal volume is called inspiratory reserve volume and the total breathing capacity of the lungs as measured by the volume which can be completely expired after maximum filling is the vital capacity Included in this is an amount called expiratory reserve volume which can be expired from the functional residual volume Total expiration might completely empty the sim ple lungs of some amphibians However as the lungs of animals elaborate with alveolar devel opment it is not possible to expire all lung con tents The remainder after limit of vital capacity is reached is called residual air and its presence at metamorphosis hatching or birth always indi cates that breathing has started The various venti lation volumes and pressures have not been meas ured in most species Some representative values are given in Table 4

Dead space Buccopharyngeal ventilation con tinues in reptiles but this mechanism no longer provides for pulmonary ventilation It remains an important adjunct to breathing however because it serves to reduce the dead space volume Dead space average normal volume for an adult human for example is about 150 ml and tidal volume about 500 ml which means an actual ventilatory volume of about 350 ml In a few air breathing animals which occupy an aquatic habitat respiration is ap-parently supplemented by buccopharyngeal breath ing of water Such a process has been described for a few species of turtles which have an especially vascular pharynx but whether an important amount of O is thus derived has not yet been as-certained This ventilation with water may actually subserve olfactory and gustatory senses

Thoracoabdominal breathing Pulmonary venti lation in reptiles utilizes a more familiar process than buccopharyngeal breathing that is it depends upon an aspiration or sucking inspiration such as in birds and man This involves development of a movable rib basket and elaboration of the inter costal musculature which by enlargement of the body cavity produces on inspiration a negative pressure in the lungs (intrapulmonary) with refer ence to the atmosphere The abdominal muscles and myoelastic tissue of the lungs and air sacs remain as in amphibians important in expiration but these are augmented by striated muscular mem branes which form diaphragms and also ensheath the lungs in some turtles

Turtles are exceptional among reptiles and air breathers generally because the ribs are fused into a shell which prohibits expansion of the body wall for inspiration In these animals muscular mem branes which enclose the viscera and others which form diaphragms at the leg pockets in the shell produce expiratory and inspiratory force respec tively

Reptiles In reptiles generally when at re t breathing cycles occur in groups which are inter spersed among long intervals of apnea During apnea the glottis is closed the lung air is under a few mm mercury positive pressure and buccopha ryngeal ventilation waxes and wanes to a degree associated at least in part with the extent of sen sory disturbance This buccopharyngeal activity re-sembles olfactory sniffing as seen in dogs The reptilian cycle of pulmonary inspiration and expi ration is much simpler than the frog s because the glottis and nares remain open while breathing movements occur and air moves freely between lung and atmosphere in direct response to the ac tion of breathing muscles in trunk and viscera These muscles act alternately to enlarge and re-duce the body cavity exerting pressure changes through tissue fluids directly on the lung air itself Inspiration is clearly by suction This is in contrast to the injection action of the buccopharyngeal mus-cles in amphibians Like amphibians however the breathing remains periodic and the glottis is nor mally closed

Birds and mammals Birds and mammals utilize strictly thoracoabdominal breathing to ventilate both nasal and pharyngeal cavities and lungs The same structures used in reptiles continue to oper ate namely trunk musculature and diaphragms The glottis does not close the lung from atmosphere normally however nasal valves are absent and buccopharyngeal movements cease Because the lung and airway are greatly elaborated over th

in previously considered classes of animals, cer tain central regions develop more critical and spe cific control than glottal valves, over air flow and diffusion. Such control regions include bronchioles and more distal sphincters at alveoli in mammals and, although less well understood, at correspond ing parabronchi and air capillaries in birds. Also other discrete structures develop as bellows, to change breathing forces into ventilation pressures. These structures are the extrapulmonary air sacs in birds (air sacs also occur in some reptiles) and the distal ducts and alveolar sacs in mammals. The alveolus becomes more strictly a diffusion exchange unit. As a result of all these specializations, great stability of physical conditions at diffusion sur faces is achieved in a system which, at the same time, accommodates high and variable rates of ex change.

The structural differences between the respira tory units of bird and mammal are subordinate to these common functional characteristics. The alve olus of mammals is usually a terminal membranous sac with a porelike orifice (the postmortem diame ter is about 70 μ in the rabbit) into an alveolar sac; occasionally they are appendant on a respira [illegible]

[illegible] diameter about 50 μ in the chicken) into a tu bular parabronchus. These alveolar structures both provide, in the aggregate, a reservoir with tremen dous surface area (at least 50 m² in man) which sequesters a mechanically stable intraalveolar at [illegible]

[illegible] pressure gradients established, as in all verte [illegible]

[illegible] tion among millions of alveoli is often called al [illegible]

[illegible] of ventilation and diffusion in respiratory exchange of man are shown in Tables 3–5.

Avian respiration. Ventilation in birds utilizes the expandable thoracic cage formed by ribs and sternum, the thoracic and abdominal muscles, and a muscular pulmonary diaphragm to produce breathing forces. Inspiration is produced largely by thoracic muscles, especially the triangular which increase the capacity *of the body cavity* mostly by outward displacement of the sternum rather than by lateral *expansion*. Expiration is produced by elastic recoil of stretched tissues and contraction of appropriate thoracic and abdominal muscles, supplemented by a pulmonary diaphragm which compresses thoracic air sacs on contraction. This pulmonary diaphragm is not homologous with that of mammals; it does not separate thorax and abdomen or contract on inspiration. Breathing *movements in birds* are usually synchronous with wing movements during flight, but there is uncer tainty about whether inspiration occurs on *the up* or down stroke of the wings. Such synchrony is also observed in flying mammals, such as the bat. Some air sacs in birds, although a minor part of the total capacity, lie among wing muscles and also form *spaces in some of the long bones. The former* are directly compressed by flight movements to aid ventilation, but the latter are removed from direct effect by any of the breathing muscles.

Air sac. Air sac walls do not accommodate ac tual *respiratory exchange. In addition to their es* sential bellows action, however, air sacs are impor tant *for enlarged respiratory capacity, for heat* elimination (particularly during flight) and to some extent for buoyancy because they occupy more than 20% of the volume of the body. The lungs themselves are much smaller and more com pact than in mammals, and they are attached ex cept ventrally to the thoracic wall and ribs; con sequently no *intrapleural space intervenes between* lung and body wall as in mammals. However, the volume changes in birds' lungs are very small. Be cause the thoracic and abdominal cavities are not separated by a diaphragm, pressure changes are *transmitted throughout,* and the *major air sacs* ac commodate as much as 75% of the tidal volume (Table 5).

Table 5. Combined measurements of respiratory values in pigeon, duck, and chicken to compare role of various sacs and lung*

	Volume % of total	Composition	
		O_2%	CO_2%
Tidal air	10 15		
Inspired	10–15	21 0	0 03 0 04
Expired	10–15	13 5	5 1–6 5
Interclavicular	20–25	14 6	5 0–6 9
Abdominal	50–70	19 0	1 9–2 7
Lung	10–12	20 0†	1 0†

* From H. H. Dukes, *Physiology of Domestic Animals* 1955; A. Krogh, *Comparative Physiology of Respiratory Mechanisms* 1941; C. L. Prosser, *Comparative Animal Physiology* 1950; P. D. Sturkie, *Avian Physiology* 1954.
† Estimated for parabronchi.

Air movement. The movement of air in and out of the bird's respiratory system is probably tidal only in certain regions of the larger airways, namely from nostrils through primary bronchi at one extreme and from air sacs through latero bronchi at the other. There is *considerable,* al though inconclusive, evidence that air movement in *the dorsoparaventrobronchi* is circulatory and uni directional. This flow is thought to result from aero dynamic characteristics of the passages, for no valves have ever been found, and breathing pres sures vary quite uniformly in all sacs. Sac pressures in the pigeon, for example, range about 6 mm of H_2O below and above atmospheric on inspiration and expiration throughout. Recurrent bronchi pro vide an accessory path for part of the incoming air *to pass through the lung.*

Buccopharyngeal breathing Buccopharyngeal breathing is indirect, when compared with thoraco abdominal It involves two distinct stages ventila tion of the mouth and ventilation of the lungs The necessary pressure gradients between mouth and atmosphere and between mouth and lungs are gen erated by muscles which raise and lower the hyoid apparatus and floor of mouth and throat, the same mechanism is used for water breathing in fish Pul monary inspiration in buccopharyngeal breathing is more descriptively an injection or adspiration The volume of air inspired or expired per breathing cycle is called tidal volume

In the frog movement of air between mouth and atmosphere requires only about 3–5 mm of water (H_2O) pressure, that between mouth and lungs reaches 25–35 mm H_2O pressure during the peak of pulmonary inspiration A gradient is directed to ward the lung of about 20 mm on inspiration be cause a volume of gas (functional residual volume) remains in the lung from the preceding expiration under about 10 mm H_2O pressure The glottis closes at the end of inspiration and a positive intrapulmonary pressure persists during apnea of about 20 mm H_2O as a result of elastic recoil and muscular tonus in lung itself and in body wall These forces also cause expiration when the glottis opens Whether buccal ventilation alone or pul monary ventilation or a combination of these oc curs depends upon neuroregulatory processes which determine the relationship of nasal and glottal valves with each other and with the breathing mus culature Thus a frog s lungs and body can be dis tended greatly beyond normal dimensions by suc cessive inspirations alone

Vital capacity The excess capacity of any ani mal to inspire beyond normal tidal volume is called inspiratory reserve volume and the total breathing capacity of the lungs as measured by the volume which can be completely expired after maximum filling is the vital capacity Included in this is an amount called expiratory reserve volume which can be expired from the functional residual volume Total expiration might completely empty the sim ple lungs of some amphibians However as the lungs of animals elaborate with alveolar devel opment it is not possible to expire all lung con tents The remainder after limit of vital capacity is reached is called residual air and its presence at metamorphosis hatching or birth always indi cates that breathing has started The various venti lation volumes and pressures have not been meas ured in most species Some representative values are given in Table 4

Dead space Buccopharyngeal ventilation con tinues in reptiles but this mechanism no longer provides for pulmonary ventilation It remains an important adjunct to breathing however because it serves to reduce the dead space volume Dead space average normal volume for an adult human for example is about 150 ml and tidal volume about 500 ml which means an actual ventilatory volume of about 350 ml In a few air breathing animals which occupy an aquatic habitat respiration is ap parently supplemented by buccopharyngeal breath ing of water Such a process has been de cribed for a few species of turtles which have an especially vascular pharynx but whether an important amount of O_2 is thus derived has not yet been as certained This ventilation with water may actually subserve olfactory and gustatory senses

Thoracoabdominal breathing Pulmonary venti lation in reptiles utilizes a more familiar process than buccopharyngeal breathing, that is, it depends upon an aspiration or sucking inspiration such as in birds and man This involves development of a movable rib basket and elaboration of the inter costal musculature which by enlargement of the body cavity produces on inspiration a negative pressure in the lungs (intrapulmonary) with refer ence to the atmosphere The abdominal muscles and myoelastic tissue of the lungs and air sacs remain as in amphibians important in expiration but these are augmented by striated muscular mem branes which form diaphragms and also ensheath the lungs in some turtles

Turtles are exceptional among reptiles and air breathers generally because the ribs are fused into a shell which prohibits expansion of the body wall for inspiration In these animals muscular mem branes which enclose the viscera and others which form diaphragms at the leg pockets in the shell produce expiratory and inspiratory force respec tively

Reptiles In reptiles generally when at rest breathing cycles occur in groups which are inter spersed among long intervals of apnea During apnea the glottis is closed the lung air is under a few mm mercury positive pressure and buccopha ryngeal ventilation waxes and wanes to a degree associated at least in part with the extent of sen sory disturbance This buccopharyngeal activity re sembles olfactory sniffing as seen in dogs The reptilian cycle of pulmonary inspiration and expi ration is much simpler than the frog s because the glottis and nares remain open while breathing movements occur and air moves freely between lung and atmosphere in direct response to the ac tion of breathing muscles in trunk and viscera These muscles act alternately to enlarge and re duce the body cavity exerting pressure changes through tissue fluids directly on the lung air itself Inspiration is clearly by suction This is in contrast to the injection action of the buccopharyngeal mus cles in amphibians Like amphibians however the breathing remains periodic and the glottis is nor mally closed

Birds and mammals Birds and mammals utilize strictly thoracoabdominal breathing to ventilate both nasal and pharyngeal cavities and lungs The same structures used in reptiles continue to oper ate namely trunk musculature and diaphragms The glottis does not close the lung from atmosphere normally however nasal valves are ab ent and buccopharyngeal movements cease Becau e the lung and airway are greatly elaborated over those

instruments which recorded abdominal movements in a horse during one breathing cycle, along with concomitant pressure changes in the designated areas. The shapes of such graphs vary in detail among different species of mammals, as can be expected from a comparison of some ventilation parameters for a few animals (Table 4). The forces of respiratory exchange are apparent from such measurements as those in Table 3, made on air and blood entering, within, and leaving the lungs of man

Vertebrate respiratory regulation. Ventilation and circulation through the lungs are controlled and adjusted to maintain an efficient exchange of gases in relationship to the needs of the animal and to the suitability of the environment Sensory units (ten ion receptors) for pressure or distension occur in lung and breathing elements, and in related circulatory organs. These signal the medulla oblongata of the brain, from which the breathing rhythm itself originates, as well as other parts of the brain, and breathing is adjusted to physiological changes and physical conditions. Other receptors, especially in certain major blood vessels near the heart and in the medulla itself, signal the respiratory areas of the brain about blood chemical conditions, in particular, about oxygen carbon dioxide, and acidity Signals from other parts of the body, such as the olfactory and tracheal membranes and the joints, and from other parts of the brain, such as heat regulating, olfactory, and phonation areas, also reach respiratory areas. Here the correlating and integrating role of the brain determines how breathing will proceed whether to speed up stop, cough, or whatever is appropriate Oxygen tension in blood is a basic regulatory ele-

factor in blood acidity, ventilation also comes to play a major role in critical regulation of hydrogen ion concentration (blood acidity) in higher animals See CAROTID BODY

Parn

... the evolutionary scale This reflects the increasing energy requirements of such biological advantages as rapid sustained locomotion and critical continuous neuroregulatory activity Through such attributes as these, animals gain increasing independence from environmental limitations Because there is about thirty times less oxygen in an aquatic environment volume for vol

... respiration In amphibians, the skin is free from such protective structures as scales, corneum, and hair, thus diffusion is much less restricted than in other animals, and the skin is an important respiratory organ in water as well as in air Permeability to respiratory gases includes permeability to water, however, and thus restricts even terrestrial amphibians such as toads to a very humid atmosphere. Some aquatic amphibians, such as the mud puppy, *Necturus*, possess lungs as well as gills, but no such animals are capable of the high energy output which characterizes each of the strictly terrestrial and air breathing vertebrate classes: the reptiles, birds, and mammals Even these classes do include some species which are partly and some strictly aquatic in habitat. All the aquatic species remain strictly air breathing, but they have modified respiratory and circulatory processes which accommodate diving and underwater activity during long suspension of breathing

the apnea which characterizes normal periodic breathing of reptiles on land In diving birds and mammals however, the suspension is a departure from the eupneic pattern, for eupnea is normally interrupted only at rest by an occasional deep breath and a pause of about 5 sec duration, such as the sigh in man A dive of more than 2-3 min duration in birds and mammals requires special adaptations, because no animal has means of storing oxygen to last more than a few minutes; however seals may submerge for 15–25 min and whales for an hour or more Important means of submergence in such animals include (1) restriction of blood flow from the bulk of the muscles, thus sequestering lactic acid and building up a debt for oxygen to be paid on access to air, (2) higher tolerance for carbon dioxide, which is always toxic in excess, and which must accumulate until breathing is resumed, (3) conservation of movement while submerged, (4) reflexes around nostrils which stop breathing on contact with water, and (5) great reduction in lung volume when the high pressure of great depths is involved, which reduces diffusion area of the lung The last of these does not concern respiration itself, rather it serves to limit nitrogen build up in tissues with its attendant hazard of bubble formation, or bends Such bubbles may form on too rapid ascent from depths, and they can cause immobilization and death Also, lungs and air sacs may serve as buoyancy organs for swimming and diving, because their contents can be altered in volume to change the displacement of the animal

Adaptations in terrestrial animals The adaptations for diving in birds and mammals have recently been found in counterpart among some nonaquatic animals such as rabbits, which may become immobile and suppress breathing under certain conditions, for example in evading predators For active locomotion in air and for highest brain activity, however oxygen is the most urgently needed of all substances required from the environment; therefore essentially continuous ventilation is a necessity for strictly terrestrial animals such as the chicken, cat, and man Such animals survive without actual ventilation only before the time of

Respiration activity Although gas pressure is uniform, gas composition differs among the air sacs, depending upon their location along the air way Respiratory measurements from the pigeon, chicken, and duck show the range of participation by sacs and lungs in respiratory activity Two of the five major groups of sacs are presented in Table 5 Such figures indicate that the tidal volume at rest is about equal to the volume of the lung, the large capacity and special distribution arrangements for air at the abdominal sacs provide for constant gas composition at alveoli nearly equal to that of the atmosphere, and the anterior sacs sequester vitiated expiratory air from dilution of the inspired air such as occurs in a strictly reciprocatory (tidal) breathing process

The lungs of birds are clearly specialized as distinct respiratory diffusion organs, even morphologically segregated from the ventilation mechanism which subserves them The over all result is that the birds' respiratory exchange structures, the air capillaries, are open to air of nearly atmospheric composition whether during inspiration or expiration The development of high magnitude and fine control of diffusion exchange in the bird's respiratory system is consonant with the very high metabolic requirements of flight

Mammalian respiration. Ventilation in mammals is more nearly like that in reptiles than in birds In mammals alveoli are terminal sacculations in a serial arrangement of ventilatory ducts and sacs rather than appendant tubules along a parallel duct arrangement as in birds A major organ of ventilation found only in mammals is the muscular diaphragm which divides thorax and abdomen Because it lies as a dome with its convex face toward the thorax when it contracts and flattens it augments thoracic muscles in increasing the capacity of the thoracic cavity The abdominal wall relaxes at the same time to accommodate viscera which are displaced by the diaphragm As a result of these movements air under pressure differential of a few millimeters of mercury (for example −1 5 mm Hg in the nasal cavity of the horse) passes into the pulmonary system The abdominal muscles cannot contribute directly to inspiration in mammals or any other vertebrates except turtles in which a special arrangement is associated with the shell as previously mentioned

Breathing mechanism It is in the nature of muscular membranes that they must be oriented as the mammalian diaphragm is, that is they must insert along their periphery into a resistant structure and bulge into the cavity, if contraction is to produce negative pressure in the cavity A positive expiratory force is a different matter, however and abdominal muscles contribute such force in all vertebrates The extent varies with the degree of ventilation—a much greater contribution in exercise than at rest—with the species and even with the sex In the human female during eupneic breathing, for example, costal movement predominates and expiration results largely from the passive recoil of lung and chest, whereas in male there is somewhat more abdominal involvement In mammals generally, abdominal muscles are more involved in expiration than is true for man, and in the larger quadrupeds the work of displacing heavy pendant viscera requires their continuous activity Forced breathing (hyperpnea) with increased amplitude, as during exercise, labored breathing (dyspnea), as during strenuous exercise or at high altitude, and compressatory acts all require the abdominal muscles These supplement the internal intercostal muscles which pull the ribs to resting position On the other hand they are little involved in panting (polypnea) or sniffing

Myoelastic fibers Myoelastic fibers of the lung itself not only provide a passive component of expiration but they also account for collapse of the lung if the chest cavity is opened to the atmosphere Because from the first filling at birth these fibers are stretched during the entire life of the animal the surface of the lung always tends to recoil from adjacent structures This recoil is limited in birds because the lung is structurally attached, but in mammals and all other animals the lung surface is free from attachment The visceral pleura of mammals, a covering membrane, adheres to the lungs, but this is in turn separated from all adjacent structures, which are covered by a parietal pleura only by a film of mucoid fluid Measurement of the elastic pulling force exerted by the lung against the surface tension of this fluid reveals the equivalent of a pressure which is negative with respect to atmosphere when the animal is at rest or during eupnea It is negative with respect to intrapulmonic pressure at all times This is called intrapleural pressure (see Table 4)

Pressure change and ventilation The relationships between various pressure changes and ventilation activity in mammalian breathing are illustrated in Fig 23 Here are shown tracings from

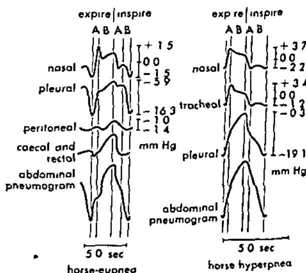


Fig 23 Composite diagram of factors in breathing cycle of horse, analyzed to show correlation of intracoelomic and intrapassage pressures with external musculature

instruments which recorded abdominal movements in a horse during one breathing cycle, along with concomitant pressure changes in the designated areas. The shapes of such graphs vary in detail among different species of mammals as can be expected from a comparison of some ventilation parameters for a few animals (Table 4) The forces of respiratory exchange are apparent from such measurements as those in Table 3 made on air and blood entering within and leaving the lungs of man.

Vertebrate respiratory regulation Ventilation and circulation through the lungs are controlled and adjusted to maintain an efficient exchange of gases in relationship to the needs of the animal and to the suitability of the environment Sensory units (tension receptors) for pressure or distension occur in lung and breathing elements, and in related circulatory organs. These signal the medulla oblongata of the brain from which the breathing rhythm itself originates as well as other parts of the brain and breathing is adjusted to physiological changes and physical conditions. Other receptors, especially *in certain major blood vessels* near the heart and in the medulla itself signal the respiratory areas of the brain about blood chemical conditions, in particular about oxygen carbon dioxide and acidity Signals from other parts of the body *such as* the olfactory and tracheal membranes and the joints, and from other parts of the brain, such as heat regulating olfactory and phonation areas, also reach respiratory areas Here the correlating and integrating role of the brain determines *how breathing* will proceed whether to speed up, stop cough or whatever is appropriate Oxygen tension in blood is a *basic regulatory element* for ventilation rate in lower vertebrates but it becomes secondary to carbon dioxide in birds and mammals. Because carbon dioxide is a major factor in blood acidity ventilation also comes to play a major role in *critical regulation* of hydrogen ion concentration (blood acidity) in higher animals. See CAROTID BODY

Resp[illegible]

[illegible]

[illegible]s the increasing energy requirements of such biological advantages as rapid sustained locomotion and critical continuous neuroregulatory activity Through such attributes as these animals gain increasing independence from environmental limitations. Because there is about thirty times less oxygen in an aquatic environment [illegible]

[illegible]

[illegible] protective structures as scales corneum and hair thus diffusion is much less restricted than in other animals and the skin is an important respiratory organ in water as well as in air Permeability to respiratory gases includes permeability to water, however and this restricts even terrestrial amphibians such as toads to a very humid atmosphere Some aquatic amphibians, such as the mud puppy, *Necturus*, possess lungs as well as gills *but no such animals are* capable of the high energy output which characterizes each of the strictly terrestrial and air breathing vertebrate classes the reptiles birds and mammals Even these classes do include some species which are partly and some strictly aquatic *in habitat All the* aquatic species remain strictly air breathing but they have modified respiratory and circulatory processes which accommodate diving and underwater activity during long suspension of breathing

Suspension of breathing In aquatic reptiles such as snapping turtles and alligators breathing suspension *for dives is simply a special instance of* the apnea which characterizes normal periodic breathing of reptiles on land In diving birds and mammals however the suspension is a departure from the eupneic pattern for eupnea is normally interrupted only at rest by an occasional deep breath and a pause of about 5 sec duration such *as the sigh in man* A *dive of more than 2 3 min* duration in birds and mammals requires special adaptations because no animal has means of storing oxygen to last more than a few minutes, however seals may submerge for 15–25 min and whales for an hour or more *Important* means of submergence in such animals include (1) restriction of blood flow from the bulk of the muscles, thus sequestering lactic acid and building up a debt for oxygen to be paid on access to air (2) higher tolerance for carbon dioxide which is always toxic in excess and which must accumulate until breathing *is resumed* (3) *conservation of movement* while submerged (4) reflexes around nostrils which stop breathing on contact with water, and (5) great reduction in lung volume when the high pressure of great depths is involved which reduces diffusion area of the lung The last of these does not concern respiration itself rather it serves to limit nitrogen build up in tissues with its attendant hazard of bubble formation or bends Such bubbles may form on too rapid ascent from depths and they can cause immobilization and death Also lungs and air sacs may serve as buoyancy organs for swimming and diving because their contents can be altered in volume to change the displacement of the animal

Adaptations in terrestrial animals The adaptations for diving in birds and mammals have recently *been found in counterpart among some* nonaquatic animals such as rabbits which may become immobile and suppress breathing under certain conditions for example in evading predators For active locomotion in air and for highest brain activity, however *oxygen is the most urgently* needed of all substances required from the environment, therefore essentially continuous ventilation is a necessity for strictly terrestrial animals such as the chicken cat and man Such animals survive with*out actual ventilation only before* the time of

birth, or hatching, although nonrespiratory periodic breathing movements do occur in egg or uterus as appropriate structures develop Early respiration utilizes special diffusion respiratory structures such as the allantois and placenta In birds an actual pulmonary ventilation, utilizing the air chamber of the egg, begins shortly before hatching In mammals it can begin only when the fetus gains direct access to air at birth See PLACENTATION, RESPIRATORY SYSTEM DISORDERS [F H MC]

Bibliography M E Brown (ed), *The Physiology of Fishes*, 2 vols, 1957, J F Fulton (ed), *Textbook of Physiology*, 17th ed, 1955, J C B Grant, *A Method of Anatomy*, 6th ed, 1952, A Keith, *Human Embryology and Morphology*, 6th ed, 1948, A A Maximow and W Bloom, *A Textbook of Histology*, 7th ed, 1957

Respiratory system disorders

The function of the respiratory system is to supply the body with the oxygen needed for metabolic activities in the cells and to remove carbon dioxide which is a product of such activity For this purpose the right ventricle of the heart pumps venous blood, containing carbon dioxide, through the pulmonary arteries into the lung capillaries, where gaseous exchanges between the blood and air occur The arterial or oxygenated blood is then drained through the pulmonary veins into the left auricle The inspiratory movements of the chest wall and diaphragm suck air through the nose, throat larynx, trachea, and bronchi into the lung alveoli, where it comes into close contact with the blood circulating in the capillaries During expiration this air is expelled through the same pathways See RESPIRATORY SYSTEM

Any disease of the respiratory system airways or blood circulation is therefore apt to reduce the respiratory function It is well known however that less than one third of the lung is needed to sustain life at rest The other two thirds more or less represent reserve volume which is used when more oxygen is needed as when the body is working Slight reduction of the respiratory function hence manifests itself in shortness of breath, dyspnea at work Dyspnea at rest is a sign of severe disease of the respiratory system In extreme conditions this is accompanied by incomplete oxygenation of the blood in the lung The blood cannot get rid of all carbon dioxide and cannot take up enough oxygen It becomes darker and has a bluish appearance where it shines through the skin as on the lips or nails This condition is called cyanosis If the respiratory insufficiency continues to increase basic functions of the organism are disturbed by the lack of oxygen The patient loses consciousness since the brain is most sensitive to lack of oxygen and finally dies if no relief can be given

Diseases of the respiratory system can be located in the lung tissue itself, interfering directly with the gas exchange, in the airways thus reducing the air brought into the lung, or in the blood vessels disturbing the blood circulation through the lung Under rare conditions the respiratory musculature can be paralyzed, as in poliomyelitis, or injured by an accident

Diseases of the lung. Changes in the content and distribution of air result from lung diseases In pulmonary emphysema a focal reduction of respiratory tissue is found with extreme enlargement of the air spaces In other diseases the air is displaced by fluid as in edema, by inflammatory exudates as in pneumonia by tissue which results from scars or cancer, or by collapse of the lung in atelectasis All these conditions can be visualized by x ray Air being more translucent than tissue every accumulation of fluid or tissue within the lung appears as a density or shadow on the x ray film or on the screen of the fluoroscope See LUNG DISORDERS, RADIOLOGY

Air-passage diseases. Most diseases of the airways increase the resistance against which air is sucked in and pushed out In order to avoid reduction of the air volume breathed, more work has to be applied to respiration The effect of a disease of the airways on respiratory function, however, depends largely on its localization Diseases of the nose have little influence since collateral respiration through the mouth compensates easily Diseases of the throat larynx and trachea have an enormous effect since the inflow of air to both lungs is rendered more difficult The same is true for a generalized narrowing of all smaller bronchi as is found in bronchial asthma, while localized narrowing of even a large bronchus will manifest itself as dyspnea at work if at all A more common sign of bronchial disease is cough due to irritation of the mucous membrane See BRONCHUS, LARYNX NOSE DISORDERS

Pulmonary circulatory diseases. Among the diseases of pulmonary circulation congenital malformations of the heart and pulmonary artery account for many cases of respiratory insufficiency of newborn and younger children called blue babies because of their cyanosis (see CARDIOVASCULAR SYSTEM) In adults acquired heart diseases such as narrowing of the mitral valve between left auricle and ventricle greatly influence the pulmonary circulation (see CIRCULATION DISORDERS) In this so called mitral stenosis the venous outflow from the lung has to be forced against the considerably increased resistance of the narrowed valve This can be overcome only by raising the blood pressure in the pulmonary blood vessels (pulmonary hypertension) The opposite is the case in narrowing of the main pulmonary artery Smaller arteries can become occluded by embolism the result of a thrombus formed in another part of the vascular system mostly in varicose veins of the thigh The thrombus is released and carried by the blood stream into the lung where it is stopped in a small artery which it occludes The region normally supplied by this vessel receives no more blood and can become an infarct (see LUNG DISORDERS) [F WL]

Bibliography W A D Anderson *Pathology* 3d ed 1957 W Boyd *Pathology for the Physician* 6th ed 1958 R L Cecil and R F Loeb (eds) *Textbook of Medicine*, 2 vols 10th ed 1959

Response

A quantitative expression of the manner in which a microphone amplifier, loudspeaker, or other component or system performs its intended function A linear response means that the output signal is exactly proportional to the input signal for the entire range of frequencies over which the device is intended to operate A logarithmic response means that the output signal is a logarithmic function of the input signal The response of a device is often presented as a curve on a graph indicating deviation over the frequency range from the response at some selected frequency, such as 1000 cycles per second An example is the frequency response curve of an amplifier *See* AMPLITUDE CHARACTERISTIC CURVE. [J MR.]

Rest mass

A constant associated with a material body which determines its inertial properties and its internal energy content. *It is sometimes called the inertial mass.*

For a particle of rest mass m_o, Newton's second law of motion is

$$m_o\, dv/dt = \mathbf{F}$$

where dv/dt is the acceleration of the particle and F is the force applied to it. For a particle moving with a speed near that of light this equation must be replaced by the relativistic equation of motion

$$\frac{d}{dt}\left(\frac{m_o \mathbf{v}}{\sqrt{1-(v^2/c^2)}}\right) = \mathbf{F}$$

The formula connecting internal energy E_o and rest mass (Einstein's mass-energy relation) is

$$E_o = m_o c^2$$

See INERTIA OF ENERGY, RELATIVISTIC MECHANICS RELATIVITY [E L HI]

Resultant of forces

A system of at most a single force and a single couple whose external effects on a *rigid body are* identical with the effects of the several actual forces that act on the body For analytic purposes

...g the method of Fig 1c to a three dimensional vector polygon in which one force is drawn from the tip of the previous one until all are laid out. The resultant force is the force vector that is required to close the polygon directed from the tail of the first force vector to the tip of the last one

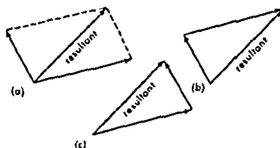
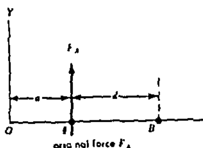


Fig 1 Resultant of two forces acting through a common center (a) Diagonal of parallelogram (b c) Hypotenuse of triangle



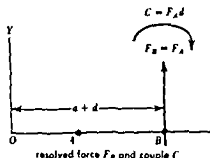


Fig 2 Resolution of a force into a force and a couple

A force system has a zero force resultant if its *vector polygon closes*

*The resultant of force along an axis may be de*sired In that case all forces are resolved into their respective components along the axis and the components are added algebraically to obtain the *corresponding resultants In this respect the resolution of a force* into components is an inverse operation to the composition of multiple forces into a resultant However whereas the resultant force is unique to a given force system a force can be *resolved into any variety of components*

Two force systems are equivalent if their resultant forces as described above are equal and if their total vector moments about the same point are also equal *Vector moments are combined in the same manner as forces* that is *by parallelo*grams triangles or polygons An alternate test for equivalence is that the total moments about two different points are respectively equal

In Fig 2 the force F_A acting at point A may be resolved into an equivalent system *consisting of* equal parallel F_B acting at point B and couple C, the magnitude of C is the moment of F_A about B

Also the total moment of F_B and C about A is the moment of F_A about A, namely zero

A resultant is the equivalent force system having the fewest possible forces and couples The resultant of concurrent forces is a force equal to their vector sum and acting through the point of concurrence A special case is a collinear system in which the resultant is collinear with the forces of the system as well as equal to their vector sum

When all forces of a system are coplanar, the resultant may be a force or a couple If it is a force the resultant is positioned to produce the same moment about a reference point as the system Should the vector sum of forces be zero the resultant is a couple that develops the same moment as the system See COUPLE

In a three dimensional force system the resultant consists of a force element passing through an arbitrary point and equal to the total force of the system and a couple element that produces a moment equal to the total moment of the system about any point on the line of action of the force element See STATICS [N S F]

Retaining wall

A wall designed to maintain differences in ground elevations by holding back a bank of material Sometimes a retaining wall also serves as a foundation wall

Material that exerts pressure on the back of the wall is called backfill It includes material taken from the excavation and replaced behind the wall The load applied on backfill above wall level is called surcharge

Earth pressure against the back of a wall is known as active pressure and acts at an oblique angle the precise angle depends upon the character of the backfill the slope of the ground surface presence of surcharge and ground water level Earth pressure against the front of a wall is known as passive pressure and may be of greater unit intensity than active pressure although the soil must be somewhat compressed before this force develops

External stability is achieved when a wall is proportioned so that it will neither rotate nor slide under all dead load and applied forces In addition to earth pressure the following forces must be considered

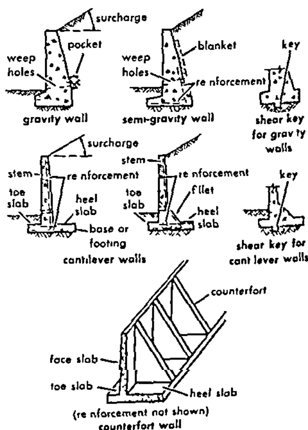
The weight of the wall acts in a vertical direction through its center of gravity Surcharge loads may consist of inclined embankments and any live load they carry for instance trucks or cranes Lateral forces may be caused by ice thrust or frost action from repeated freezing and thawing of poorly drained soil If ice layers can form behind a wall pressures may be larger than a wall can be designed to resist Thrust action from clay soils may result from repeated changes in water content for when the soil drys it cracks and the cracks fill with loose soil leaving too little space for expansion in place when the next wetting occurs Earthquakes may cause lateral forces Vibrations from machinery or traffic may increase the effects of earth pressure

Uplift under the footing will occur if the water level is above that point or if water is not drained from behind the wall Uplift may also occur in bed rock if it is seamed or slightly porous Retaining walls that also serve as foundation walls or abutments may in addition experience lateral forces from cranes wind, building framing or tractive effort Bridge spans cause vertical longitudinal and transverse lateral forces on retaining walls used as abutments See BRIDGE

Vertical resisting forces are supplied by reactions of the soil or rock under the footing and are known as bearing pressures Horizontal resisting forces are supplied by friction under the footing or by shear keys extending below the footing Passive pressure in front of the wall is sometimes omitted from stability analyses because of the uncertainty as to its magnitude and the amount of movement necessary to bring it into action

Drainage A retaining wall is drained primarily to prevent accumulation of water in the backfill thus avoiding hydrostatic pressure, formation of ice that may cause thrust swelling of cohesive backfills and decrease in the stability of the soil Highly permeable material such as gravel or crushed stone should be located in the backfill to collect ground water Surface water also should be drained but its entrance into the backfill should be minimized by use of relatively impervious topsoil or paving

Water should be led away through weep holes built into the wall or through corrosion resistant pipe having open joints or perforations Pocket



The major types of retaining wall

drains, a pocket of stone or gravel at weep holes, may be used if the backfill is moderately permeable. Blanket drains consist of a layer of stone or gravel against the entire back surface of the wall. Water discharge should be carried away by gravity or pumped from collection sumps.

Types of wall. The illustration shows the major types of retaining wall. Gravity walls are of massive, solid construction proportioned so that ten

and partial disintegration is not as serious since stability depends on weight. Gravity walls are usually low. Unit stresses in the concrete are very low.

Semigravity walls are constructed with narrower stems. A few tension bars are built into the back and toe. Considerable concrete is saved by using small amounts of steel.

A cantilever wall is formed of three cantilever beams: the stem, toe projection, and heel projection. Reinforcing steel is required in all members. The simplest form is used for low walls. For higher walls, a fillet may be an economical further reinforcement. Sliding resistance may be improved by a wall design with a key projecting downward into the soil. This type is the most common. Members can be so designed that concrete and steel unit stresses equal critical values of these materials.

A counterfort wall is a thin reinforced concrete-face slab backed up by deep vertical cantilever stems, or counterforts. The heel slab of the wall footing carries the backfill load horizontally to the counterforts. The wall may be constructed with a key to increase sliding resistance. The counterfort wall is often the most economical type of high retaining wall. See FOUNDATIONS. [R.D.C.]

Bibliography. See FOUNDATIONS.

Reticular formation (brain)

...reticular formation is thought to be a complex, highly integrated mechanism which exerts some degree of inhibition or facilitation on almost every type of activity of the nervous system.

Anatomy. The brain stem, composed of the medulla oblongata, the pons, and the midbrain or mesencephalon, is the basic integrating and connecting unit of the central nervous system. The long ascending sensory fibers pass upward from all parts of the body and are distributed largely in the brain stem to ...

...ters, such as the spinal cord. In addition the brain stem is itself the site of many important structures, notably the cell clusters or nuclei of cranial nerves. See BRAIN; NERVOUS SYSTEM.

The reticular formation lies in and around these other more definite structures which, by their presence or passage, break up the reticular formation into many small islands of gray matter connected by large numbers of relatively short nerve fibers. These fibers pass in an apparently haphazard manner in all directions, but recent work reveals that this reticulum is in reality a highly complex, intricately organized master system of communication which alters many body activities.

Many authorities disagree on the exact anatomic or physiologic extent of the reticular formation. This is due partly to differences in methods of study from one discipline to another. Although certain reticular formation cell clusters are well circumscribed, others are not. This fact, in addition to the multiplicity of connections that are hard to trace and because of the proximity to other better-known structures, has prevented a more detailed analysis to date.

Influence on behavior. For many years the reticular formation was largely ignored except for a vague descriptive acknowledgement of its presence. Recently, however, increased attention has focused on this portion of the brain, mainly because of increasing evidence of its vital, though often subtle, role in many body activities. Sleep, wakefulness, attention, and other aspects of consciousness, as well as effects on muscular coordination, vascular tone, blood pressure, and many other aspects of everyday adaptation to environment, are all affected by the condition and responses of the reticular system. See SLEEP.

The appearance of drugs such as chlorpromazine which seem to affect selectively the reticular structures has created an upsurge of interest and investigation in which many of the techniques of pharmacology have been added to the methods of anatomy and physiology to gain further insight in this field. See ANATOMY, REGIONAL; PHYSIOLOGY, GENERAL; TRANQUILIZER.

The use of electrical stimulation and electroencephalography have been added to the microscopic study of neurons, fibers, and degeneration phenomena following experimental injury to selected sites. See ELECTROENCEPHALOGRAPHY.

Refinements in techniques have led to observations which support the general view that the reticular formation may exert a facilitory or inhibitory action on at least three major areas of nervous activity.

One area of activity so affected is the alteration of ascending impulses received from almost any sensory receptor in the body. The passage of such impulses apparently stimulates the reticular system through a collateral system. The level of activation depends upon previous correlation between cortical areas and the reticular structures involved, so that a kind of presetting mechanism is involved. An illustration may be helpful. A person goes to bed with the conscious or unconscious knowledge that waking will be signaled by the noise of an alarm clock. Despite many other stimuli, some of a high degree of intensity, the person sleeps soundly until the ring of the alarm. Of

Also, the total moment of F_B and C about A is the moment of F_A about A, namely zero

A resultant is the equivalent force system having the fewest possible forces and couples The resultant of concurrent forces is a force equal to their vector sum and acting through the point of concurrence A special case is a collinear system in which the resultant is collinear with the forces of the system as well as equal to their vector sum

When all forces of a system are coplanar, the resultant may be a force or a couple If it is a force the resultant is positioned to produce the same moment about a reference point as the system Should the vector sum of forces be zero, the resultant is a couple that develops the same moment as the system *See* COUPLE

In a three dimensional force system, the resultant consists of a force element passing through an arbitrary point and equal to the total force of the system and a couple element that produces a moment equal to the total moment of the system about any point on the line of action of the force element *See* STATICS [N S F]

Retaining wall

A wall designed to maintain differences in ground elevations by holding back a bank of material Sometimes a retaining wall also serves as a foundation wall

Material that exerts pressure on the back of the wall is called backfill It includes material taken from the excavation and replaced behind the wall The load applied on backfill above wall level is called surcharge

Earth pressure against the back of a wall is known as active pressure and acts at an oblique angle, the precise angle depends upon the character of the backfill the slope of the ground surface presence of surcharge and ground water level Earth pressure against the front of a wall is known as passive pressure and may be of greater unit intensity than active pressure although the soil must be somewhat compressed before this force develops

External stability is achieved when a wall is proportioned so that it will neither rotate nor slide under all dead load and applied forces In addition to earth pressure the following forces must be considered

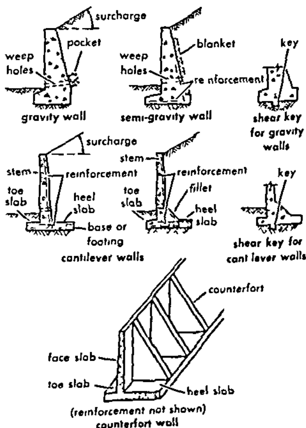
The weight of the wall acts in a vertical direction through its center of gravity Surcharge loads may consist of inclined embankments and any live load they carry for instance trucks or cranes Lateral forces may be caused by ice thrust or frost action from repeated freezing and thawing of poorly drained soil If ice layers can form behind a wall pressures may be larger than a wall can be designed to resist Thrust action from clay soils may result from repeated changes in water content for when the soil drys it cracks and the cracks fill with loose soil leaving too little space for expansion in place when the next wetting occurs Earthquakes may cause lateral forces Vibrations from machinery or traffic may increase the effects of earth pressure

Uplift under the footing will occur if the water level is above that point or if water is not drained from behind the wall Uplift may also occur in bed rock if it is seamed or slightly porous Retaining walls that also serve as foundation walls or abutments may in addition experience lateral forces from cranes, wind, building framing or tractive effort Bridge spans cause vertical, longitudinal and transverse lateral forces on retaining walls used as abutments *See* BRIDGE

Vertical resisting forces are supplied by reactions of the soil or rock under the footing, and are known as bearing pressures Horizontal resisting forces are supplied by friction under the footing or by shear keys extending below the footing Passive pressure in front of the wall is sometimes omitted from stability analyses because of the uncertainty as to its magnitude and the amount of movement necessary to bring it into action

Drainage A retaining wall is drained primarily to prevent accumulation of water in the backfill thus avoiding hydrostatic pressure, formation of ice that may cause thrust swelling of cohesive backfills and decrease in the stability of the soil Highly permeable material such as gravel or crushed stone should be located in the backfill to collect ground water Surface water also should be drained but its entrance into the backfill should be minimized by use of relatively impervious topsoil or paving

Water should be led away through weep holes built into the wall or through corrosion resistant pipe having open joints or perforations Pocket



The major types of retaining wall

picule reticulations *Titusidia* from the Missis- ... pian is an example *See* HEXACTINELLIDA, HEX... ATEROPHORA [W D H]

Retinitis

An inflammatory condition involving the retina the light-sensitive innermost coat of the eye In the majority of cases the adjacent middle layer of the eyeball (the choroid layer which contains the blood vessels of the eyeball) is also involved in which case the condition is known as chorioretinitis The combination of all forms of retinitis and chorio-

...

here to include the degenerative retinopathies which are associated with a variety of generalized chronic diseases.

Reaction to injury The various components of the retina and the choroid react to injury in different ways The specialized light receptors (rods and cones) and the nerve cells of the retina have no capacity for regeneration and when injured either

...

scars which can grow into and replace damaged regions of the *retina* The formation of a scar either fibrous or glial results in permanent impairment of vision The affected portion of the retina is rendered insensitive to light causing a defect in the visual field known as a blind spot or scotoma The amount of visual impairment depends *on the* location as well as the size of the *defect* thus a small defect in the region of central distinct vision is more serious than a *much larger one near the* periphery of the retina

Etiology Retinitis may be due to infectious agents (usually bacteria less commonly fungi rarely viruses) mechanical injury of the eyeball (contusion) or intense light (photoretinitis) Chronic progressive damage to the retina (degenerative retinopathy) is frequently associated with conditions such as diabetes high blood pressure arteriosclerosis, senility leukemia and anemia

Infections of the retina and choroid are usually secondary occurring as a result of *direct extension* from adjacent infected tissues (cornea iris nasal cavities, eyelids) or via the bloodstream from distant infected organs (kidney heart valves lungs) Contusion of the retina ... loss of v...

... Photoretinitis may follow observation of intense light sources such as *the sun electrical arcs, welding operations* or even reflected sun (snow blindness) The resulting blind spot is central and is usually temporary, but it may be permanent if exposure is severe enough

The degenerative retinopathies although differing slightly depending on the condition with which they are associated have as a common feature abnormalities of the blood vessels of the retina, *which result in impairment of blood flow recurrent*

...

disease characterized by growth of pigment cells into and over the inner surface of the retina and resulting in ultimate scarring and blindness *See* EYE, EYE DISORDERS [W R A D]

Retrograde motion (astronomy)

In astronomy either an apparent east-to-west motion of a planet or comet with respect to the background stars or a real east-to-west orbital motion of a comet about the Sun or of a satellite about its primary The majority of the objects in the solar system revolve from west to east about their primaries *However near the time of closest approach* of Earth and a superior planet such as Jupiter because of their relative motion the superior planet appears to move from east to west with respect to the background stars The same apparent motion occurs for an inferior planet such as Venus near the time of closest approach to Earth

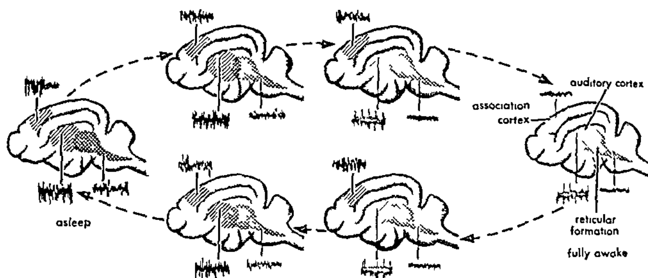
Actual rather than apparent retrograde motion *occurs among the satellites and comets the eighth* and ninth satellites of Jupiter and the ninth satellite of Saturn are examples [R L D]

Reverberation

After sound has been produced in or enters an enclosed space it will be reflected repeatedly by the boundaries of the enclosure even after the *source ceases to emit sound This prolongation of sound after the original source has stopped is* called reverberation A certain amount of reverberation adds a pleasing characteristic to the acoustical qualities of a room However excessive reverberation can ruin the acoustical properties of an otherwise well designed room A typical record representing the sound pressure level at a given point in a room plotted against time after a sound source has been turned off is given in the decay curve shown in Fig 1 The rate of sound decay is *not uniform but fluctuates about an average slope*

Reverberation time Because of the importance of the proper control of reverberation in rooms a standard of measure called reverberation time (abbreviated t_{60}) has been established Reverberation time is the time required for sound to die away to one-thousandth of its initial pressure that is to drop 60 decibels (db) in sound pressure level

Optimum reverberation time is a matter of individual preference A critical study of empirical data based upon preference evaluations in the *United States and abroad has been made by V O*



Sleeping and waking appear to result from the interaction of opposed processes in the brain The alert pattern (gray) begins in the reticular formation, whence it spreads to other brain areas The sleep pattern (black) appears early in the association cortex and gradually encompasses the rest of the brain (From V Rowland, Conditioning and brain waves Sci American 201(2) 89–96, 1959)

ten, in fact, the alarm need not actually ring because the warning to wakefulness may be elicited by the preliminary click of the alarm control

This illustrates at least two principles involved in the modification of ascending activating impulses first, the inhibitory or damping effect of the reticular formation on most sensory impulses, and second the primary importance of some level of cortical association and control over the reticular alarm system

The same type of reaction, and many others have been repeatedly demonstrated in animal experiments wherein electroencephalographic records can be correlated with cortical patterns in many states relating to consciousness

The various tranquilizers are thought to work through a similar mechanism to produce an inhibition of ascending stimuli and also a decrease in reverberation within the brain itself

The second area in which nervous activity is altered by the reticular formation is in its relations with the cerebellum Investigation in this field is not yet extensive enough to be conclusive but certain observations can be made Under particular conditions the reticular formation may facilitate or inhibit certain actions involving motor movement and coordination Somewhat conflicting observations have been reported in regard to specific movements or specific animals yet it is apparent that many cerebellar activities are dependent upon intact nuclei in the reticular substance notably the paramedian reticular nuclei the lateral reticular nuclei, and others

Much of this relationship is that of a complex feedback system, wherein continuing motor activity requires continuing adjustment, or coordination of both sensory and motor impulses by the modifying reticular structures The inhibition of extensor tone in decerebrate animals is a gross example of one effect of this relationship Many more subtle influences particularly those involving the localization of motor functions between specific sites in the cerebellum and reticular nuclei, have been reported No concise explanation can be given until further studies are made but the implications are of great importance in terms of the so called primitive reflex acts and other motor activities.

The third major area to be influenced by the reticular formation is that of impulses mediated by the spinal cord Facilitation or inhibition of cortical or reflex movements by reticular elements have been repeatedly demonstrated In addition, general and specific effects on vasomotor tone on muscular tonus and on both the inspiratory and expiratory phases of respiration have been traced to alterations induced by reticular activity or lack of it

Summary. It may be useful to visualize the reticular formation as a complex highly integrated mechanism which exerts some degree of inhibition or facilitation on almost every type of nerve body activity The dualistic response cannot be overemphasized particularly because the cortically arranged selectivity of the inhibition facilitation is a primary feature The mediation of sensory, motor, and integrative impulses touches on specific body activities as well as on the more significant bodily states of wakefulness sleep attention and related conditions of whole body activity Finally coordination and reflex activity require reticular formation participation Thus the reticular formation emerges from obscurity to become a fascinating regulatory mechanism at present only dimly perceived *See* PSYCHOLOGY PHYSIOLOGICAL AND EXPERIMENTAL, REFLEX CONDITIONED [F C S T]

Reticulosa

An order of the subclass Hexasterophora in the class Hexactinellida This is a group of Paleozoic hexactinellids with a branching form Each branch is provided with dermal parenchymal and gastral

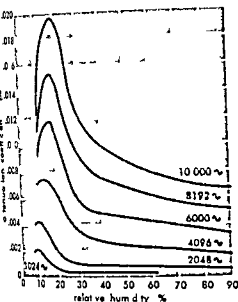


Fig. 4. Values of the attenuation coefficient m as a function of relative humidity for different frequencies. (After V. O. Knudsen and C. M. Harris, *Acoustical Designing in Architecture*, Wiley, 1950)

For frequencies above 2000 cps, especially in large auditoriums, the effects of air absorption must be included in the reverberation time formulas. The corresponding equations are then

$$t_{60} = \frac{0.049V}{S[-2.30\log_{10}(1-\bar{a}) + 4mV]} \text{ sec}$$

and $$t_{60} = \frac{0.049V}{S\bar{a} + 4mV} \text{ sec}$$

where m is the attenuation coefficient given in Fig. 4. It can be shown that air absorption is molecular in origin. *See* ARCHITECTURAL ACOUSTICS; SOUND.

[C.M.H.]

Bibliography. V. O. Knudsen, C. M. Harris, *Acoustical Designing in Architecture*, 1950.

Revetment

A means of protecting river banks against bank erosion. In some cases the revetment must act as a solid barrier; however, as a basic principle it should be designed to induce the high velocity thread of the current to move away from the threatened area. This is most readily accomplished when (1) the revetment is initiated at a point upstream from the point of attack and at an angle of not more than about 15° to the current and (2) the surface of the revetment is rough.

The rough surface generates a zone of turbulence which, in effect, acts as a cushion. Rapidly fluctuating pressures accompany the turbulence and tend to leach underlying material unless an effective filter is supplied. This filter can be provided by proper dumping of quarry run rock or by placing a blanket of gravel under the surface material of the revetment.

Alignment of revetments should be smooth, without holes or projections, and sharp breaks in alignment should be avoided, since they may cause the current to cross abruptly to attack the opposite bank or may cause eddy action along the bank to scour or undermine the revetment.

Revetments may consist of the following types: (1) rock paving with a heavy rock toe base for underwater protection, (2) rock, asphaltic, or concrete paving with brush, lumber, or concrete mattress for the underwater portion, (3) pile dikes with rock or mattress (brush or lumber) for underwater protection.

On streams such as the Missouri and Arkansas rivers, where the depth of water at low water levels seldom exceeds 15–25 ft, the use of pile dikes and rock paving is effective and economical. Quarry run rock is used, and its size must be such that at least 30% of the rock is large enough to resist movement by the current. Usually quarry run rock is well graded, and placement by dumping provides both an inverse filter and the desired rough surface.

On deeper channels such as the Mississippi, where depths may exceed 100 ft, articulated concrete or reinforced asphaltic mattresses are generally used.

Pile dikes or other permeable fence-type revetments placed parallel to the bank generate local

are generally effective if the angle of attack of the current does not exceed 30°. For permanence, particularly in streams with heavy ice or drift runs



Stone revetment



Pile-dike revetment

Knudsen and C M Harris (Fig 2) Since the optimum reverberation time for music depends on the type of music it is represented in the figure by a broad band The optimum reverberation time for a room used primarily for speech is considerably shorter, reverberation times longer than those shown for speech result in a decrease in speech intelligibility The optimum reverberation time at frequencies other than 500 cycles per second (cps) is obtained by multiplying the 500 cps value by the ratio R which is given in Fig 3 Note that R is unity for frequencies above 500 cps and is given by a band for frequencies below 500 cps For large rooms R may have any value within the indicated band for small rooms preferred ratios are in the lower part of the band

Mean free path According to the principles of geometrical acoustics sound radiated from a source in an enclosure is successively reflected by its boundaries The average distance between reflections is defined as the mean free path The mean free path of a sound ray in a room depends on the shape and size of the room and to some extent on the distribution and nature of the absorptive material However in most cases it is approximately $4V/S$ ft where V is the volume of the room and S the total surface area

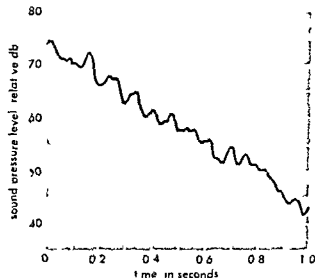


Fig 1 Typical decay curve illustrating reverberation

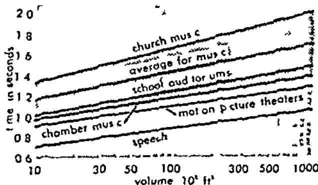


Fig 2 Optimum reverberation time at 500 cps for different types of rooms as a function of room volume This figure should be used in conjunction with Fig 3 to obtain optimum reverberation time as a function of frequency (After V O Knudsen and C M Harris Acoustical Designing in Architecture Wiley 1950)

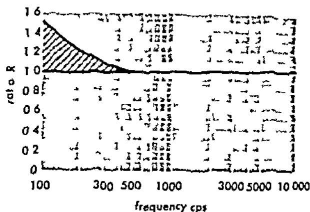


Fig 3 Chart for computing optimum reverberation time as a function of frequency The time at any frequency is given in terms of the ratio R which should be multiplied by the optimum time at 500 cps (from Fig 2) to obtain the optimum time at that frequency (After V O Knudsen and C M Harris Acoustical Designing in Architecture, Wiley 1950)

Decay rate The number of reflections per second of a decaying sound wave is numerically equal to the distance sound will travel in 1 sec that is, the velocity of sound c which is about 1130 ft/sec in air at 20°C divided by the average distance between reflections or the mean free path Hence the number of reflections per second is $cS/4V$ Each time a wave strikes one of the boundaries, on the average a fraction ($\bar{\alpha}$) of the energy is absorbed and a fraction $(1-\bar{\alpha})$ is reflected where $\bar{\alpha}$ is the average absorption coefficient given by

$$\bar{\alpha} = \frac{\alpha_1 S_1 + \alpha_2 S_2 + \alpha_3 S_3 +}{S_1 + S_2 + S_3 +}$$

where α_1 is the coefficient of absorption of surface S_1 and so forth Because sound pressure is proportional to the square root of sound intensity the ratio of the average reflected pressure to incident pressure is given by $(1-\bar{\alpha})^{1/2}$ so that the average decrease in the sound pressure level is

$$10 \log_{10} \frac{1}{1-\alpha} \quad \text{db/reflection}$$

Since there are $cS/4V$ reflections per second the average decay rate is

$$1230\, S/V[-2.30 \log_{10}(1-\bar{\alpha})] \text{ db/sec}$$

Reverberation time formulas From the preceding equation for decay rate it follows that the time it takes for the sound pressure level to decay 60 db that is the reverberation time is

$$t_{60} = \frac{0.049V}{S[-2.30 \log_{10}(1-\bar{\alpha})]} \text{ sec}$$

When $\bar{\alpha} \ll 1$ this equation becomes

$$t_{60} = \frac{0.049V}{S\bar{\alpha}} \text{ sec}$$

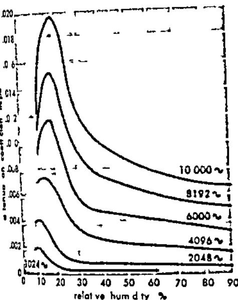


Fig. 4. Values of the attenuat on coeff c ent *m* as a function of elat ve humid ty for d fferent frequenc es. After V O Knudsen and C M. Harr s Acoust cal Designing in Architecture W ley 1950)

For freq encies above 2000 cps especially in large auditoriums, the effects of air absorption must be in luded in the reverberation time formulas The corresponding equations are then

$$t_{60} = \frac{0.049V}{S[-2.30 \log_{10}(1-\bar{a}) + 4mV]} \quad \text{sec}$$

and
$$t_{60} = \frac{0.049V}{S\bar{a} + 4mV} \quad \text{sec}$$

e *m* is the attenuation coefficient given in Fig can be shown that air absorption is molecular rigin. See ARCHITECTURAL ACOUSTICS SOUND [C M H]

bliography V O Knudsen C M Harris *ust cal Designing in Architecture* 1950

vetment

neans of protecting river banks against bank ero- n. In some cases the revetment must act as a ld barr er however as a basic principle it ould be des gned to induce the high velocity read of the current to move away from the threat ed a ea. This is most readily accomplished when 1) the e etment is initiated at a point upstream rom the point of attack and at an angle of not no e than about 15° to the current, and (2) the urface of the revetment is rough

The rough surface generates a zone of turbulence which, n effect, acts as a cushion Rapidly fluctuat ng p es ures accompany the turbulence and tend to leach underl ing material unless an effective filter is uppl ed. This filter can be provided by proper dump ng of quarry run rock or by placing a blanket of gra el under the surface material of the revet ment.

current to cross abruptly to attack the opposite bank or may cause eddy action al ng the bank to scour or undermine the revetment

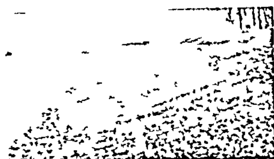
Revetments may consist of the following types (1) rock paving with a heavy rock toe base for un derwater protection (2) rock asphaltic or con crete paving with brush lumber or concrete mat tress for the underwater portion (3) pile dikes with rock or mattress (brush or lumber) f r under water protection

On streams such as the Missouri and Arkansas rivers, where the depth of water at low water levels seld m exceeds 15 25 ft the use of pile dikes and rock paving is effective and eco nomical Quarry run rock is used and its size must be such that at least 30% of the rock is large enough to resist movement by the current Usually quarry run rock is well graded and placement by dumping provides both an inverse filter an l the desired rough surface

On deeper channels such as the Mississippi where depths may exceed 100 ft articulated con crete or reinforced asphaltic mattresses are gener ally used

Pile dikes or other permeable fence-type revet ments placed parallel to the bank generate local

are generally effective if the angle of attack of the current does not exceed 30° For permanence par ticularly in streams with heavy ice or drift runs



Stone revetment



P le-d ke revetment

they should be protected with rock to a height which causes enough accretion to support willow growth

Groins are short dikes, approximately perpendicular to the bank, which are sometimes used in lieu of revetment, particularly in areas where the bed is relatively stable They are usually of solid construction, either sheet piling or dumped materials For other information on control of rivers *see* RIVER ENGINEERING [D C B , W E J]

Reynolds number

In fluid mechanics, the ratio $\rho v d/\mu$ of the inertia force $\rho v d$ to the viscous force μ, where ρ is fluid density v is velocity d is a characteristic length, and μ is fluid viscosity The Reynolds number is significant in the design of a model of any system in which the effect of viscosity is important in controlling the velocities or the flow pattern (*see* MODEL THEORY) In the evaluation of drag on a body submerged in a fluid and moving with respect to the fluid the Reynolds number is important If the model is operated in the same fluid as is the prototype the similarity requirement based on the Reynolds number yields

$$v_m = nv$$

It is evident that if both gravity and viscosity are significant and if the same gravitational field and same fluid are used in both the model and prototype similarity must be achieved for Froude and for Reynolds numbers (*see* FROUDE NUMBER) The only possible solution is that n be equal to unity or a model be equal in size to the prototype If the length scale is to be greater than 1 (model smaller than the prototype), either the model must be tested in a different gravitational field or a different fluid must be used to satisfy similitude requirements *See* DYNAMIC SIMILARITY

The Reynolds number also serves as a criterion of type of fluid motion In a pipe for example laminar flow normally exists at Reynolds numbers less than 2000 and turbulent flow above about 3000 *See* PIPE FLOW, SHIP PROPULSION [C M]

Rhabdiasoidea

A group of parasitic nematodes set apart by their characteristic morphology and by their life cycles which may include a parasitic generation alternating with one or more free living generations This group may be given the rank of an order or superfamily by specialists One species causes an important disease of man, others parasitize domestic animals and cause great economic loss, still others are of great biological interest because of their varied methods of reproduction and compli

children being most often infected In the parasitic generation there are no male worms The females live in the intestinal wall and produce eggs parthenogenetically The larvae hatching from the eggs in the intestinal lumen and passing in the feces, are known as first stage larvae When deposited on soil under favorable conditions of temperature and moisture, they may undergo so-called

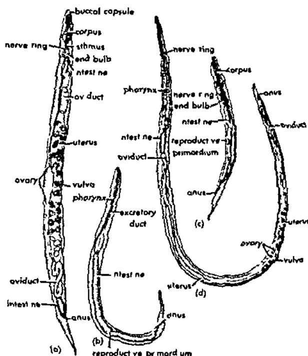


Fig 1 Rhabdiasoidea *Strongyloides ransomi* (a) Free-living female, 1 mm long (b) Filariform young (c) Rhabditiform young (d) Parasitic female, 4 mm long (After B Schwartz and J Alicata 1930, from L H Hyman, The Invertebrates vol 3, McGraw-Hill, 1951)

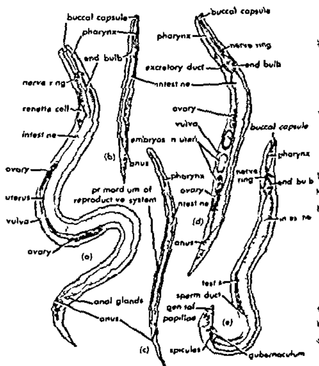


Fig 2 Rhabdiasoidea *Rhabdias buloni* (a) Parasitic female (b) Rhabditiform young (c) Filariform young (d) Free-living female (e) Free-living male (After E Metschnikoff, 1865 from L H Hyman The Invertebrates, vol 3, McGraw-Hill, 1951)

direct development into an infective larval stage in less than a week. At other times they develop "indirectly" into adult males and females of a free-living generation. After fertilization these females produce eggs from which first-stage larvae hatch. These larvae may develop directly or indirectly. In the latter case the pattern may be repeated many times. There are both genetic and environmental factors determining developmental patterns.

The infective larvae arising by either method can penetrate intact human skin. In the skin they migrate into blood vessels and are carried by the blood through the heart to the lungs. There they lodge in the capillaries, then penetrate into the air sacs, *migrate up the trachea, are swallowed, and* complete their development in the intestine.

Other species. Other species in the genus *Strongyloides* (Fig. 1) have slightly different life cycles. Species of other genera, notably *Rhabdias (Fig. 2)* in amphibians and reptiles, have even more complicated cycles. In some cases the parasitic females are protandrous, that is, they produce sperm early in life which fertilize eggs produced in a later female phase of the same worm. [J.A.S.]

Bibliography. A. C. Chandler, *Introduction to Parasitology*, 9th ed., 1955; L. H. Hyman, *The Invertebrates*, vol. 3, 1951.

Rhabditoidea

Small to moderately sized nematodes with the esophagus of two bulbs and with small pitlike amphids. Most nematologists consider this group to be a superfamily, although some give it the status of an order. Mainly terrestrial, they range in habits from saprophagous to parasitic. The life cycle is direct, but sometimes includes an infective larva that must be transported by an invertebrate. The larva is ensheathed in the cuticle of a previous stage which prevents desiccation. Most species live in decaying plant and animal matter and utilize saprophytic insects for transportation. In some species the nematode becomes a passive endoparasite, but at least one family contains active parasites of insects, and another family parasites of snails. Certain genera are always found in soil about plant roots and may have some *parasitic relationship*. Members of another family are adapted to unusual habitats, such as vinegar and beer mats. See NEMATODA.

[H.E.W.]

Rhabdocoela

Formerly considered an order of the Turbellaria, the Rhabdocoela is now usually divided into the three orders, Catenulida, Macrostomida, *and* Neorhabdocoela. The rhabdocoeles have a simple unbranched intestine, with little if any diverticulations. The pharynx is simple or bulbous, the gonads are usually compact, a cuticular *apparatus is as*sociated with the copulatory organ, and there are two main longitudinal nerves. This large group includes nearly all of the small fresh-water Turbellaria as well as many marine and some *terrestrial* species.

Catenulida. This group comprises threadlike, colorless fresh-water rhabdocoeles with a simple pharynx and a single median protonephridium. *They lack yolk glands and are usually without sex* organs. Reproduction is largely asexual by binary fission with the formation of chains of zooids. When present the female system consists entirely of one or more ovaries without ducts or gonopore. After fertilization the entolecithal eggs reach the exterior through rupture of the body wall. The testis is single and has a duct enlarged at its end to form an unarmed copulatory organ which opens to the exterior through an anterior dorsal male genital pore. Ciliated pits or grooves and sometimes statocysts *are present.* Eyes are usually lacking although light-refracting bodies which may serve as photoreceptors are present in some species. The nervous system has cerebral ganglia and four pairs of *longitudinal nerves. The genus Stenostomum is the* commonest and most widely distributed of all fresh-water Turbellaria and was the subject of some of C. M. Child's studies on axial gradients.

Macrostomida. *These* rhabdocoeles have a simple pharynx, paired protonephridia, and a single pair of longitudinal nerves. They are without yolk glands but have oviducts. There is generally a single pair of compact testes. From each testis a vas deferens extends posteriorly, often enlarging to form a spermiducal vesicle. Usually the male system also includes a seminal vesicle, prostate glands and vesicle, and a copulatory organ armed with a tubular cuticular stylet. The male genital pore is posterior to the female on the midventral surface.

There are two families, Microstomidae and Macrostomidae. The Microstomidae are elongate cylindric in shape and seldom have sex organs present since reproduction is largely asexual by fission with the formation of chains of zooids. Ovary and oviduct when present are single. In the genus *Microstomum* nematocysts obtained through the ingestion of *Hydra* are often present in the epidermis and may be utilized for offense and defense as in *Hydra*. The Macrostomidae are broad and flattened in shape and often have a spatulate posterior end which is adhesive. No asexual reproduction occurs. *Ovary and oviduct are nearly always paired;* eggs are entolecithal, and accessory structures such as a bursa or vagina may be present. The male system includes testes, sperm ducts, a seminal vesi*cle, prostate glands and vesicle, and a penis armed* with a cuticular stylet.

Neorhabdocoela. The neorhabdocoeles are fresh-water, marine, or terrestrial rhabdocoeles with a *bulbous pharynx, paired protonephridia, sexual* reproduction, yolk glands combined with or separate from the ovaries, ectolecithal eggs, and ventral gonopores. This group includes most of the genera *and species of rhabdocoeles and is divided into* three sections: the Dalyelloida, Typhloplanoida, and Kalyptorhynchia.

Dalyelloida. These neorhabdocoeles have the *mouth at or near the anterior end,* a cask-shaped pharynx, no proboscis or rhammite tracts, and a

they should be protected with rock to a height which causes enough accretion to support willow growth.

Groins are short dikes, approximately perpendicular to the bank, which are sometimes used in lieu of revetment, particularly in areas where the bed is relatively stable. They are usually of solid construction, either sheet piling or dumped materials. For other information on control of rivers *see* RIVER ENGINEERING [D.C.B., W.E.J.]

Reynolds number

In fluid mechanics, the ratio $\rho v d/\mu$ of the inertia force $\rho v d$ to the viscous force μ, where ρ is fluid density, v is velocity, d is a characteristic length, and μ is fluid viscosity. The Reynolds number is significant in the design of a model of any system in which the effect of viscosity is important in controlling the velocities or the flow pattern (*see* MODEL THEORY). In the evaluation of drag on a body submerged in a fluid and moving with respect to the fluid, the Reynolds number is important. If the model is operated in the same fluid as is the prototype, the similarity requirement based on the Reynolds number yields

$$t_a = nv$$

It is evident that if both gravity and viscosity are significant and if the same gravitational field and same fluid are used in both the model and prototype, similarity must be achieved for Froude and for Reynolds numbers (*see* FROUDE NUMBER). The only possible solution is that n be equal to unity or a model be equal in size to the prototype. If the length scale is to be greater than 1 (model smaller than the prototype), either the model must be tested in a different gravitational field or a different fluid must be used to satisfy similitude requirements. *See* DYNAMIC SIMILARITY.

The Reynolds number also serves as a criterion of type of fluid motion. In a pipe, for example, laminar flow normally exists at Reynolds numbers less than 2000 and turbulent flow above about 3000. *See* PIPE FLOW, SHIP PROPULSION. [C.M.]

Rhabdiasoidea

A group of parasitic nematodes set apart by their characteristic morphology and by their life cycles which may include a parasitic generation alternating with one or more free-living generations. This group may be given the rank of an order or superfamily by specialists. One species causes an important disease of man, others parasitize domestic animals and cause great economic loss, still others are of great biological interest because of their varied methods of reproduction and compli

children being most often infected. In the parasitic generation there are no male worms. The females live in the intestinal wall and produce eggs parthenogenetically. The larvae, hatching from the eggs in the intestinal lumen and passing in the feces, are known as first stage larvae. When deposited on soil under favorable conditions of temperature and moisture, they may undergo so-called

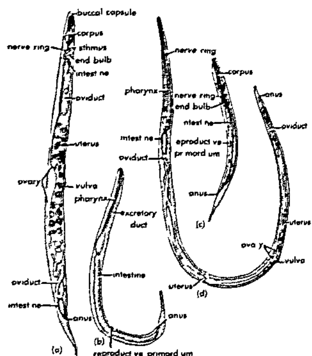


Fig. 1 Rhabdiasoidea *Strongyloides ransomi* (a) Free-living female 1 mm long (b) Filariform young (c) Rhabdiform young (d) Parasitic female, 4 mm long (*After B. Schwartz and J. Alicata 1930, from L. H. Hyman, The Invertebrates, vol. 3, McGraw-Hill, 1951*)

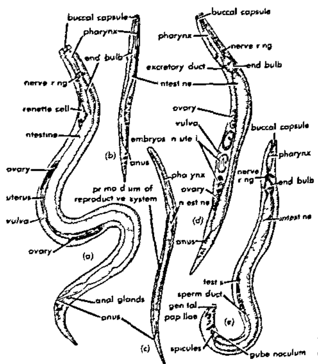


Fig. 2 Rhabdiasoidea *Rhabdias bufonis* (a) Parasitic female (b) Rhabdiform young (c) Filariform young (d) Free-living female (e) Free-living male (*After E. Metschnikoff 1865, from L. H. Hyman, The Invertebrates, vol. 3, McGraw-Hill, 1951*)

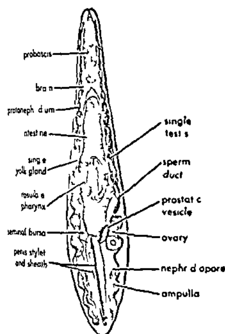


Fig. 4. *Gy atrix hermaphroditus* (From L. H. Hyman, The Inverteb ates vol 2 McGraw-H ll, 1951)

and in the absence of epidermal cilia in most species. These three modifications of the typical rhabdocoele s ructure are *correlated* with their mode of li e s nce they are ectocommensals on the bodies of larger animals. They feed on small aquatic or [illegible] and are found only in tropical and subtropical waters. *See* TURBELLARIA [E.R.J.]

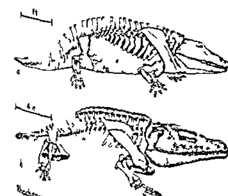
Bibliography L. von Graff, *Rhabdocoelida*, 1913

Rhachitomi

A group of labyrinthodont amphibians character

[illegible]

minating the Permian Some members of the order as *Eryops* and *Cacops*, had stout limbs and may ha e been in considerable measure terrestrial in habit, others were more purely aquatic The rhachitomous type of vertebra can be directly derived from that of crossopterygian fishes and the group may have developed at an early stage in amphibian history The rhachitomes were ancestral to the Trematosauria and Stereospondyli of the Triassic and may have been ancestral to other amphibian orders as well *See* AMPHIBIA FOSSILS LABYRINTHODONTIA [A S R]



Rhachitomous amph b ans from the Lower Perm an (a) Eryops, about 5 ft n length (b) Cacops about 16 in long (After Gregory and W ll ston)

Rhamnales

A small order of the plant subclass Dicotyledoneae, with only two families the buckthorn family (Rhamnaceae) having 45 genera and 550 species and the vine family (Vitaceae) with 11 genera and 600 species The order is characterized by a single whorl of stamens opposite the petals and the ovary surrounded by a disk The bark of *Rhamnus purshiana* yields the drug Cascara sagrada used as a mild laxative All varieties of grapes Boston ivy and Virginia creeper belong to the vine family *See* GRAPE CULTURE *see also* DICOTYLEDONEAE EMBRYOPHYTA PLANT KINGDOM [P D S]

Rheiformes

An order of birds consisting of the single family Rheidae containing the two living species of rheas The larger and better known species the common rhea (*Rhea americana*) inhabits grasslands of Brazil Uruguay and Argentina A smaller species (*Pterocnemia pennata*) is found from southeastern Peru to the Straits of Magellan Once placed in the artificial assemblage of Ratites the rheas are now thought to be related to a South American order of flying birds the Tinamiformes Rheas differ from the superficially similar ostrich in their smaller size feathered head and neck 3- instead of 2 toed feet and several other anatomical characters *See* AVES RATITES TINAMIFORMES [K C P]

Rhenium

Element [illegible] a transition element [illegible] and O B [illegible] ing success of the application of Mendeleev s periodic law The blank space in the periodic chart corresponding to dvi manganese had long occupied the attention of scientists and many efforts had been made to discover the element whose properties would entitle it to be placed in that position Noddack Tacke and Berg reasoned that it should occur in platinum ores and certain other minerals notably columbite *See* PERIODIC TABLE

Today rhenium is obtained as a by product of some metallurgical operations and is commercially available in laboratory quantities Examination of a large number of minerals showed that rhenium is widely distributed in nature The most promising source appears to be molybdenum glance MoS_2 Rhenium metal can be prepared easily by reduction of any of a number of its compounds by hydrogen It is a dense metal (21 04) with the very high melting point of 3440°C

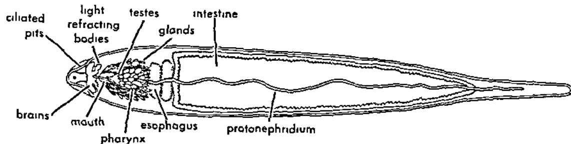


Fig 1 *Stenostomum grande* (After Nuttycombe and Waters, 1938)

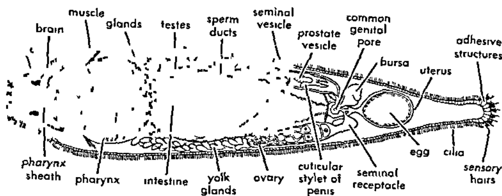


Fig 2 *Microdalyellia rossi* (After Ruebush and Hayes, 1939)

single gonopore Usually the testes are paired and the penis is armed with a cuticular stylet Ovaries and yolk glands may be combined or separated and single or paired If separate there is often a single ovary and paired yolk glands Accessory structures such as a uterus and a bursa may also occur There are several families and a number of genera and species some being commensal in other marine invertebrates while a few are parasitic Many of the commonest and most cosmopolitan fresh water species belong to the family Dalyelliidae

Typhloplanoida Neorhabdocoela with a rosulate, occasionally dolioform, pharynx, a mouth opening near the middle of the body and no proboscis There is often a common gonopore but if not, the male pore is anterior to the female pore Usually male and female gonads are paired and ovaries and yolk glands are separate In many species the glands which produce rhammites are grouped in such a manner that the rhammites form definite tracts as they move toward the anterior margin of the body The chief family is the Typhloplanidae the largest family of rhabdocoeles, which includes a great many fresh water and terrestrial forms *Mesostoma ehrenbergii*, a well known cosmopolitan species is an unusually large broad flat, and trans parent rhabdocoele It produces both summer and winter eggs The former are retained in the body until after hatching, and the young worms may often be seen squirming around in the uterus Winter eggs take much longer to develop and ap parently complete their development in the water after the death of the parent

Kalyptorhynchia These neorhabdocoeles have a rosulate pharynx a mouth opening somewhat an terior to the middle of the body, a proboscis, and lack rhammite tracts The proboscis is a protrusible muscular organ sometimes armed with hooks a[nd] richly supplied with glands It is located in a pou[ch] at the anterior end of the body and is used in ca[p]turing food In the Schizorhyncha the proboscis [is] bifurcated whereas in the more common Eukalypt[o]rhyncha it is undivided In general the members [of] this suborder are marine, living in the sand of sha[l]low waters, and most of them have been describ[ed] since 1920 The best known species is the cosm[o]politan *Gyratrix hermaphroditus*, which is comm[on] in fresh brackish and salt water

Temnocephalida. A group of rhabdocoeles whic[h] are closely related to the Dalyelloida are the temn[o]cephalids They are sometimes considered a distin[ct] order but are usually classified under the Neorhab[docoela] They differ from other rhabdocoeles chiefl[y] in the possession of tentacles and adhesive organ[s]

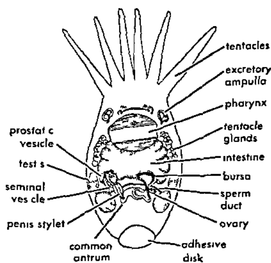


Fig 3 *Temnocephala* (After Haswell, 1893 from L H Hyman, *The Invertebrates, vol 2*, McGraw Hill, 1951)

perature of about 160°C, sulfur changes from a

rise, thus counteracting the usual decrease of viscosity of liquids with temperature rise

When linear high polymers change their state they introduce holes only in directions normal to the length of the chain. Thus a linear polymer melts in two directions but retains its solid like properties along the chain length. The result is that high polymers progress by wriggling a segment at a time as shown by the fact that they exhibit the expected pressure and temperature coefficients of a molecule the size of a segment. The viscosity is much higher for a high polymer because of a negative entropy of activation corresponding to correlation of segment motion. Extensive quantitative correlations of rheological properties are available. See POLYMER [H.E.Y.]

Bibliography F. R. Eirich (ed.), *Rheology Theory and Applications* 3 vols., 1956–1958 (vol. 3 in prep.). M. Reiner, The flow of matter, *Sci American*, 201(6) 122 138, 1959

Rheostat

the current in a circuit. The resistive element of a rheostat may be a metal wire or ribbon, carbon disks, or a conducting liquid. See RESISTOR; see also POTENTIOMETER (VARIABLE RESISTOR)

The metallic type is the most common. The wire or ribbon is constructed in a coil or a grid and taps are brought out from different sections of the element to a multicontact switch which can short circuit any desired section of the resistor or switch it out of the circuit. For more continuous control, as needed for laboratory rheostats, a sliding-contact finger bears directly on closely wound coils of resistive wire.

The carbon-disk type is used only for small currents. The resistive element is varied by changing the pressure on the carbon disks. The advantage of this type is its capacity for fine adjustment.

The electrolytic type is ideally suited to large currents. This type consists of a tank of conducting liquid in which electrodes are placed. The variation of resistance is obtained by changing the distance between the electrodes, the depth of immersion of the electrodes, or the resistivity of the solution. This type, also called water rheostat, has perfectly continuous adjustment.

Rheostats are used whenever it is desired to vary resistance or adjust current. Typical applications are for starting, controlling the speed of motors, for adjusting generator characteristics, for controlling storage battery charging, for dimming [illegible] ... [illegible] artificial loads on electric ... test.

Rheumatic fever

A childhood illness which follows an infection by *Streptococcus hemolyticus*, group A, by about 3 weeks. It is characterized by arthritis (redness and swelling of joints) and carditis (inflammation of heart tissue). Typically the polyarthritis is migratory with involvement of two or more larger joints. Other symptoms include nose bleeds and chorea, also known as St. Vitus dance. Fever, rapid pulse, paleness, and indisposition are present. Rheumatic fever is important because it can cause heart disease. The likelihood of this occurring and of permanent heart damage are increased by rheumatic fever's natural tendency to recur. Heart damage can show up later, though the active illness was decidedly mild.

Rheumatic fever, not in itself an infectious disease, follows 2–3% of untreated streptococcal infections. There is agreement as to its cause, but its mechanism is not known. The illness is not related to severity of infection nor the serologic type (Lancefield) of hemolytic streptococcus. Heredity and environment are difficult to evaluate separately as factors. Rheumatic fever does not appear after penicillin treatment of the streptococcal infection. Continuous, that is, prophylactic, use of penicillin in smaller dosage will prevent later attacks. Therapy in active rheumatic fever includes penicillin and salicylates (aspirin). Aspirin produces prompt subsidence of fever and arthritis. Bed rest is usual. Cortisone, a hormone, is administered when there is carditis. Badly scarred heart valves are corrected surgically. See LANCEFIELD DIFFERENTIATION SCHEME; SCARLET FEVER; STREPTOCOCCUS [P.L.B.]

Rheumatism

A term used to denote any combination of muscle or joint pain, stiffness, or discomfort arising from nonspecific disorders. It is generally used as a lay expression to indicate a chronic or recurrent condition affecting a certain area and precipitated by cold, dampness, or strain.

Fibromyositis refers to acute or chronic symptoms of tenderness, stiffness, and pain in a joint and related structures which follows exposure, strain, trauma, or infection. Myositis denotes an inflammation of a muscle; myalgia refers to muscle pain or tenderness without inflammation. Fibrositis is an inflammation of fibrous connective tissue, usually that of a joint region.

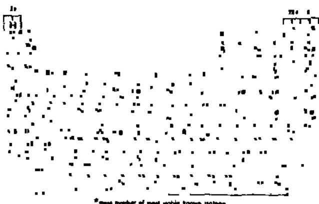
Rheumatism includes all of the above nonspecific disorders and is best reserved for complaints not related to a specific disease such as arthritis, rheumatic fever, trichinosis, or others that may cause the same or similar symptoms.

Lumbago, wryneck, charleyhorse, and shinsplint are commonly used expressions included under the catch-all category of rheumatism. [F.G.A.T.]

Rhinoceros

Any of five species of odd-toed hoofed massive mammals of the family Rhinocerotidae, order Perissodactyla, found in Africa, Asia, and the East

Rhenium is similar to its homolog technetium in that it may be oxidized at elevated temperatures by oxygen to form the volatile heptoxide, Re_2O_7, this in turn may be reduced to a lower oxide, ReO_2 The compound ReO_3, as well as several others such as Re_2O_3 and Re_2O, is well known The heptavalent oxide, Re_2O_7, may be dissolved in water to form the colorless perrhenic acid, $HReO_4$, analogous to $HClO_4$, $HMnO_4$, and $HTcO_4$, and a large series of salts corresponding to this acid have been prepared Perrhenic acid is a strong monobasic acid and is only a very weak oxidizing agent Complex perrhenates such as cobalt hexammine perrhenate $[Co(NH_3)_6(ReO_4)_3]$ are also known



The halogen compounds of rhenium are very complicated and a large series of halides and oxyhalides have been reported The halides and oxyhalides of the higher oxidation states tend to be quite volatile and sometimes are liquid Tetravalent rhenium also forms a series of double salts corresponding to hexachlororhenic acid which is analogous to those formed by technetium Organic complex salts of rhenium such as nitron perrhenate can be used in its determination

Rhenium forms two well characterized sulfides Re_2S_7 and ReS_2 as well as two selenides Re_2Se_7 and $ReSe_2$ The sulfides have their counterparts in the technetium compounds Tc_2S_7 and TcS_2 *See* TECHNETIUM, TRANSITION ELEMENTS

[S F]

Bibliography J G F Druce *Rhenium*, 1948

Rheology

The study of the deformation and flow of matter The states of matter differ strikingly in their density and in the ease with which they can be deformed The less dense the state of matter the more easily deformable it ordinarily is The viscosity of a gas arises from the crossing over of molecules from a fast moving layer into a neighboring more slowly moving layer and vice versa Because this crossing over increases with temperature rise the viscosity of a gas increases with temperature On the contrary solids and liquids become more fluid with temperature rise *See* VISCOSITY OF GASES VISCOSITY OF LIQUIDS

A perfect crystal according to theoretical calculations, should be orders of magnitude stronger than crystals customarily are found to be This is because of the presence in actual crystals of various types of imperfections which greatly facilitate their deformation The thermodynamic properties of crystals are readily calculated by assuming that the atoms form a perfectly ordered lattice The rheological properties of crystals cannot be calculated from the perfect lattice model, however, but only by considering the number and nature of the lattice imperfections Besides various types of vacant lattice sites, extra interstitial atoms are frequently present For details on the important types of crystal imperfections, *see* CRYSTAL DEFECTS

The extra fluidity of liquids as compared with solids arises from the great increase in the number of imperfections introduced with the 10% expansion in melting shown by normal liquids Ice is an exception since it contracts by about 10% on melting This contraction results from a change from the tetrahedral hydrogen bonded structure displayed by one crystalline form of ice into the close-packed structure of another form As a result, water has a viscosity of about 17 millipoises at the melting point, which is close to the normal value for ordinary liquids (A poise is a unit of viscosity equal to 1 dyne sec/cm²) A solid melts at that temperature at which the entropy from imperfections multiplied by the temperature equals the heat of introducing these imperfections

Metals melt with only 3% expansion instead of the usual 10% expansion of normal un ionized molecules because the positive ion being only about one third as large as the atom requires only one-third the space for extra equilibrium positions

Molten salts, on the other hand expand about 22% on melting In this case new equilibrium positions require about double the usual space This is probably because a sodium chloride molecule must be accommodated in the added equilibrium position instead of an Na^+ ion or a Cl^- ion separately

That the viscosity of a system closely reflects structure is exemplified by the fact that nearly all simple substances have a viscosity of the liquid at the melting point of approximately 2 centipoises Furthermore the reciprocal of the viscosity, the fluidity is a linear function of the molecular volume

Because holes are necessary to permit viscous flow and since pressure tends to decrease the number of holes in a system it follows that pressure normally increases viscosity The pressure coefficient of viscosity indicates that the empty space a molecule requires in order to flow viscously is about one seventh its normal volume To provide such vacant sites an activation energy of about one-third the heat of vaporization is required as shown by the temperature coefficient of viscosity The normal effect of pressure rise and temperature drop in increasing viscosity is modified in the rare cases where such changes promote significant modifications in structure Thus increasing the pressure on water just above the melting point increases the fluidity of water because the liquid structure is shifted away from the hydrogen bonded tetrahedral state toward the more fluid close-packed structure At a tem

perature of about 160°C, sulfur changes from a comparatively fluid, straw-colored, 8 membered ring to a dark, viscous, high polymer having thousands of atoms in a linear chain. Lubricants like wise are made relatively temperature independent by adding high polymers which are soluble only with difficulty and which unfold with temperature rise thus counteracting the usual decrease of viscosity of liquids with temperature rise

When linear high polymers change their state, they introduce holes only in directions normal to the length of the chain Thus a linear polymer melts in two directions but retains its solid like properties along the chain length. The result is that high polymers progress by wriggling a segment at a time, as shown by the fact that they exhibit the expected pressure and temperature coefficients of a molecule the size of a segment The viscosity is much higher for a high polymer because of a negative entropy of activation corresponding to correlation of segment motion. Extensive quantitative correlations of rheological properties are available. *See* POLYMER [H EY]

Bibliography F R Eirich (ed), *Rheology Theory and Applications*, 3 vols , 1956–1958 (vol 3 in prep), M Reiner, The flow of matter. *Sci American*, 201(6) 122–138, 1959

Rheostat

A variable resistor constructed so that its resistance value may be changed, without interrupting the circuit to which it is connected It is used to vary the current in a circuit The resistive element of a rheostat may be a metal wire or ribbon, carbon disks, or a conducting liquid *See* RESISTOR *see also* POTENTIOMETER (VARIABLE RESISTOR)

[illegible]

ment to a multicontact switch which can short circuit any desired section of the resistor or switch it out of the circuit For more continuous control as is needed for laboratory rheostats, a sliding contact finger bears directly on closely wound coils of resistive wire

The carbon disk type is used only for small currents The resistive element is varied by changing the pressure on the carbon disks The advantage of this type is its capacity for fine adjustment

The electrolytic type is ideally suited to large currents This type consists of a tank of conducting liquid in which electrodes are placed The variation of resistance is obtained by changing the distance between the electrodes the depth of immersion of the electrodes or the resistivity of the solution. This type also called water rheostat, has perfectly continuous adjustment

Rheostats are used whenever it is desired to vary resistance or adjust current Typical applications are for starting or controlling the speed of motors for adjusting generator characteristics, for controlling storage battery charging, for dimming lights and for imposing artificial loads on electrical equipment during test

Rheumatic fever

A childhood illness which follows an infection by Streptococcus hemolyticus, group A, by about 3 weeks It is characterized by arthritis (redness and swelling of joints) and carditis (inflammation of heart tissue). Typically the polyarthritis is migratory with involvement of two or more larger joints. Other syn [illegible]

also know [illegible]

paleness, [illegible]

fever is important because it can cause heart disease The likelihood of this occurring and of permanent heart damage are increased by rheumatic fever's natural tendency to recur. Heart damage can show up later, though the active illness was decidedly mild

Rheumatic fever, not in itself an infectious disease, follows 2–3% of untreated streptococcal infections There is agreement as to its cause, but its mechanism is not known. The illness is not related to severity of infection nor the serologic type (Lancefield) of hemolytic streptococcus Heredity and environment are difficult to evaluate separately as factors Rheumatic fever does not appear after penicillin treatment of the streptococcal infection Continuous, that is prophylactic, use of penicillin in smaller dosage will prevent later attacks Therapy in active rheumatic fever includes penicillin and salicylates (aspirin) Aspirin produces prompt subsidence of fever and arthritis Bed rest is usual Cortisone a hormone is administered when there is carditis Badly scarred heart valves are corrected surgically *See* LANCEFIELD DIFFERENTIATION SCHEME, SCARLET FEVER, STREPTOCOCCUS [P L B]

Rheumatism

A term used to denote any combination of muscle

[illegible]

tion affecting a certain area and precipitated by cold dampness or strain

Fibromyositis refers to acute or chronic symptoms of tenderness stiffness and pain in a joint and related structures which follows exposure strain trauma or infection Myositis denotes an inflammation of a muscle myalgia refers to muscle pain or tenderness without inflammation Fibrositis is an inflammation of fibrous connective tissue, usually that of a joint region

Rheumatism includes all of the above nonspecific disorders and is best reserved for complaints not related to a specific disease such as arthritis, rheumatic fever trichinosis or others that may cause the same or similar symptoms

Lumbago, wryneck, charleyhorse and shinsplint are commonly used expressions included under the catch all category of rheumatism [E G ST]

Rhinoceros

Any of five species of odd toed hoofed massive mammals of the family Rhinoceridae order Perissodactyla found in Africa Asia, and the East

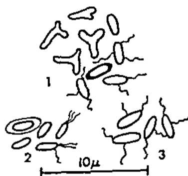


The rhinoceros, *Rhinoceros bicornis*, length to 8½ ft (From P M Duncon, ed, Cassell's Natural History, Cassell)

Indies Rhinoceroses have three toes on each foot, and unusually thick, nearly naked skins, often laid in deep folds over the body They have small ears and small eyes The African and one Asian species have two horns located on the median portion of the snout the posterior one usually smaller than the other Two of the Asian species have a single, large, anterior horn These horns are solid and are made up of a mass of cornified fibers which are modified hairs, they have no bony connection to the skull Both sexes have horns of similar size The record horn 53 in long grew on a female African rhinoceros All of the species are becoming increasingly rare and appear destined for extinction *See* PERISSODACTYLA [J D B]

Rhizobiaceae

A family of bacteria of the order Eubacteriales Some species are effective in the soil for nitrogen fixation and others are plant pathogens The bacteria comprising the three genera are heterotrophic, rod shaped gram negative and aerobic Carbohydrates are utilized by all strains without appreciable acid formation gas is not produced Isolation and cultivation of all strains within the genera



1—*Rhizobium* 2—*Agrobacterium*
3—*Chromobacterium*

Some genera of Rhizobiaceae (V B D Skerman)

Rhizobium and *Agrobacterium* are accomplished with ease, species in the genus *Chromobacterium* are difficult to maintain in culture collections

Rhizobium, the type genus of the family, comprises organisms generally known as the root nodule bacteria, or rhizobia These bacteria are short, motile rods in early stages of growth on ordinary culture media, pleomorphic, x, y, star and club-shaped bacteroidal forms are commonly found in cultures grown in acid media or in the presence of glucosides or alkaloids as well as within the nodules

The rhizobia are best known for their abilities to invade the root hairs of leguminous plants and cause the formation of cortical hypertrophies known as root nodules Free atmospheric nitrogen is fixed by the plant and bacteria in symbiosis The six species now recognized, *R meliloti*, *R trifolii*, *R leguminosarum*, *R phaseoli*, *R japonicum*, and *R lupini*, are defined by their reactions in litmus milk and their respective abilities to produce nodules on certain leguminous plants contained within cross inoculation plant groups Strains of each species show marked host specificities in their abilities to symbiose with the plants within each group The amount of nitrogen fixed by different leguminous plants varies within a wide range No correlation exists between the cultural physiological, and biochemical characteristics of the rhizobia and their abilities to symbiose with their host plants The treatment of leguminous seeds with bulk preparations of effective rhizobia is widely practiced throughout the world for the purpose of preventing nitrogen starvation lessening plant demand for soil nitrogen, and for improving the yield and quality of leguminous plants used in agriculture and soil conservation *See* NITROGEN CYCLE, RHIZOSPHERE, SOIL MICROBIOLOGY

Agrobacterium is a genus of seven species, of which five are plant pathogens *A tumefaciens* causes root and stem galls on plants contained in over 40 families of angiosperms *A gypsophilae* produces galls on *Gypsophila paniculata* L and related plants *A pseudotsugae* is pathogenic on the Douglas fir, *Pseudotsuga taxifolia* Britton *A rhizogenes* causes a hairy root condition of apple and other plants and *A rubi* is pathogenic on various species of raspberries and blackberries *A radiobacter* and *A stellulatum* are nonpathogenic, the former is commonly found on the roots of plants and as a contaminant within leguminous nodules The latter has marine habitats None of the agrobacteria fixes atmospheric nitrogen

Chromobacterium is a genus of four species of saprophytic soil and fresh water bacteria which produce moist to rugose colonies on agar media and form ring or surface membranous pellicles on gelatin and liquid media One of the best known characteristics of the chromobacteria is the production of a violet blue pigment which is soluble in alcohol, but not in water or chloroform The best known pigment type violacein is presumably related chemically to indigo

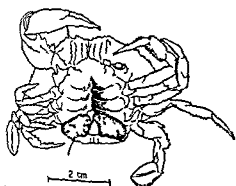
The type species is *C violaceum* Comparative

C marismortui isolated from water of the Dead Sea, a halophile with an optimum growth at about 12% sodium chloride See EUBACTERIALES [O V A]

Rhizocephala

An order of crustacean parasites related to the barnacles World wide in distribution they prey on other crustaceans, principally Decapoda such as crabs, shrimp and their allies. See DECAPODA (CRUSTACEA) In 1949 the rhizocephalan *Loxothyla*

ture females to acquire precociously the adult form These parasites have become so modified by their mode of life that, as adults, they are no longer recognizable as barnacles or even as crustaceans The clue to their relationships is found in their life history See PARASITIC CASTRATION



Loxothylacus texanus attached to the abdomen of the blue crab *Callinectes sapidus* (From E G Reinhard Parasitic castration of Callinectes 8 of Bull 98(3) 277–288 1950)

...kes, or sense organs Even an alimentary tract is missing Instead it possesses a threadlike root system which penetrates the interior of the host in all directions and absorbs nourishment from the body fluids of the crab Fertilized eggs develop in a brood chamber An opening at the summit of the sac allows sea water to enter for purposes of respiration and provides an avenue of escape for the larvae. A short stalk securely rivets the base of the sac to the abdomen of the crab.

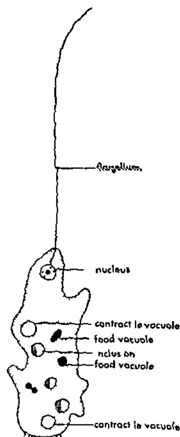
Fertilized eggs become nauplii These larval parasites, expelled from the mother s brood pouch measure 0 25 mm in length They have the appearance of barnacle nauplii except that they lack an alimentary tract. Free swimming nauplii metamorphose into cypris larvae which appear sexually indifferent. Those that settle on crabs become kentrogen larvae *A cell mass injected by the kentrogen* develops into a tumor and root system on the intestine of the host After 8 weeks the endoparasite emerges as a small sac Cypris that settle on immature Rhizocephala become larval males A cell mass injected into the brood pouch by these males migrates to the testes and differentiates into sperm Six weeks later the sac is sexually mature

Rhizocephala with the exception of the family Sylonidae, were considered hermaphroditic possessing ovaries and testes in *the same individual In* 1958 A Ichikawa and R Yanagimachi proved that the testes" in *Peltogasterella socialis were really* seminal receptacles The sperm in these female repository organs is produced exclusively by cypris male cells Adults of *Peltogasterella* are thus true females carrying hyperparasitic larval males The hermaphroditic nature of other Rhizocephala is now questionable [P C R]

Bibliography H Boschma The species of the genus *Sacculina* (Crustacea Rhizocephala) *Zool Mededeel 19(3–4)* 187 328 1937 G *Smith Rhizocephala* Fauna und Flora des Golfes von Neapel 29 1906

Rhizomastigida

An order of the class Zoomastigophorea also known as the Rhizomastigina All Rhizomastigida species are microscopic ameboid and have one or two flagella *Multicilia* has many flagellumlike proc



Mastigamoeba reptans

esses, but these may be radiating axopodia, each ending in a basal granule *Mastigamoeba* (see illustration) has a single flagellum arising from a basal granule at the nuclear membrane, with lobose pseudopodia In *Mastigella* no nuclear connection is apparent Rhizomastigida are small, colorless, and rarely abundant They generally occur at the mud water interface and are holozoic, saprozoic, or parasitic *Multicilia, Pteridomonas,* and *Actinomonas* are rounded, other species are flattened Life histories are poorly known, except for *Mastigella* The nuclear division of *Bodopsis godboldi* is typical The group is a small one, but species occur in both fresh and salt water *See* ZOOMASTIGOPHOREA [J B L]

Rhizopodea

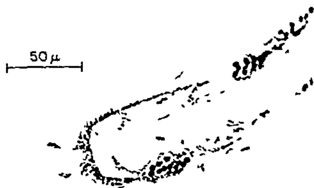
A class of the subphylum Sarcodina whose members do not have axopodia Instead, the pseudopodia may be lobopodia, filopodia, or myxopodia (rhizopodia) The Rhizopodea are often referred to as the Rhizopoda The class includes the following orders Proteomyxida, Mycetozoida, Amoebida, Testacida, and Foraminiferida *See* SARCODINA, see also articles on the orders [R P H]

Rhizosphere

The soil region subject to the influence of plant roots It is characterized by a zone of increased microbiological activity and is an example of the relationship of soil microbes to higher plants Other examples are mycorhiza (a fungus plant relationship) and bacterization (inoculation of soil or seed with microbes)

A sharp boundary cannot be drawn between the rhizosphere and the soil unaffected by the plant (edaphosphere) At the root surface the rhizosphere effect is most intense falling off sharply with increasing distance

Growth of a plant markedly changes the microbial population of soil within its influence In the rhizosphere there are more microorganisms than in soil distant from the plant This increase is most pronounced with bacteria but is evident with other groups especially actinomycetes and fungi Algae and protozoa increase less than other microorganisms The effect may be revealed by plating methods and confirmed by examination of slides buried in contact with roots, which reveal accumulations of organisms near and at the root surface. The rhizosphere effect is seen in seedling plants, it increases with the age of the plant and usually reaches a maximum at the stage of greatest vegetative growth Upon death of the plant the microbial population reverts to the level of the surrounding soil Within the range of soil moisture suited to plant growth, the number of rhizosphere organisms increases with decreasing moisture content Leguminous plants support higher rhizosphere populations than nonlegumes The stimulation of microorganism growth in the rhizosphere results chiefly from the liberation of readily available organic substances by the growing plant



Bacteria clustered about tip of live rootlet (Dr T Gibson, World Crops, 3(4), 1951)

The plant exerts a characteristic effect by shifting the balance between groups of bacteria with respect to morphological and taxonomic type, physiology, and nutritional requirements In the rhizosphere there is a preferential stimulation (increase in growth) of gram negative rods and a relative suppression (decrease in growth) of gram positive rods coccoid rods, and aerobic sporeformers Chief among those stimulated are species of *Pseudomonas, Agrobacterium,* and *Rhizobium Rhizobium* is stimulated more particularly by legumes *Arthrobacter, Azotobacter,* and nitrifying forms are less abundant in the rhizosphere than the preceding three species Rhizosphere bacteria are physiologically more active than those in nonrhizosphere soil Judged by oxygen consumption isolates from the rhizosphere show a greater degree of metabolic activity than isolates from soil regions distant from the plant In the rhizosphere, there are higher proportions of motile and chromogenic bacteria and of organisms able to decompose sugars starches, and proteins Cellulose decomposing forms are stimulated by the presence of cellular debris sloughed off by the roots Vitamin synthesizing organisms are much more numerous in the rhizosphere

From the standpoint of bacterial nutrition the most characteristic rhizosphere effect is the preferential stimulation of organisms whose food needs are met by amino acids This is due chiefly to liberation of amino acids from plant roots Bacteria depending upon the more complex nutrients and growth factors provided by soil extract (organic plant exudates) are less abundant in the rhizosphere

The rhizosphere effect is related to certain aspects of plant disease Varieties of flax tobacco, and banana susceptible to certain pathogenic fungi exert a more pronounced rhizosphere effect than resistant varieties Oat varieties susceptible to manganese deficiency disease show greater numbers of manganese-oxidizing organisms in the rhizosphere than resistant varieties Some soil amendments (additives) are found to reduce the incidence of scab in potatoes by causing an increase in the rhizosphere of organisms antagonistic to the pathogen

Mycorhiza This is a special fungus plant relationship in which fungi have a more intimate asso

cation w h roots than that shown by the general hizo phere population *Ectotrophic mycorhizal* fungi do not penetrate root cells but form closely wo en masses of hyphae about the rootlets and may enter root tissue between cells This type of myco-rhiza occurs commonly with forest trees Endo-trophic fungi penetrate more deeply *into the tis-sues and in ade the cells.* This condition is more widespread and occurs in orchids heather, fruit trees, and shrubs The significance of the associa tion is not well understood The fungi obtain food partly *from the plant. On the other hand mycorhi* zas may function as an extension of the root system in extra ting plant nutrients from soil Thus myco-rhizal fungi frequently stimulate plant growth

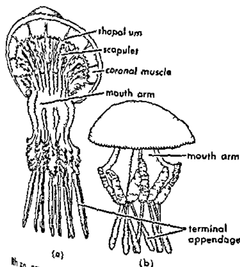
Bacterization The inoculation of soil or seed w h *microorganisms to stimulate plant growth or* to protect plants again t soil borne pathogenic or gan sms is called bacterization The value of inocu lating seed of leguminous plants with cultures of *s mbiotic nitrogen fixing* bacteria adapted to the crop is established There *is no conclusive evidence* that inoculation is of value for nonlegumes for wh h purpose cultures of the nonsymbiotic nitro-gen fixing organism *Azotobacter* have been exten s ely tr ed. *Attempts to control plant pathogenic* fungi or bacteria by heavy inoculation of soil with cul ures of antagonistic (antibiotic producing) or ganisms have not been successful under field con ditions In *natural soils* the antagonist finds itself

en ironment *to* encourage antagonists normally present in the soil. *See* BACTERIA FUNGI NITRO-GEN CYCLE SOIL MICROBIOLOGY [A G L.]

Rhizostomeae

An o der of the class Scyphozoa with the most h ghly organized features of this class The *um brella* is generally higher than it is wide except in *Cassiopea and Cephea* The margin of the umbrella is divided into many lappets but is not provided with tentacles Many radial canals which are con nected with each other to form a complicated net work issue from the cruciform *stomach The oral* part is very complicated and is 8 sided with many suctorial mouths surrounded by numerous small tentacles Usually there is no central mouth The development of eggs into medusae through the stages of planula scyphopolyp *strobila and* ephyra, *is similar* to that of the Semaeostomeae However the metamorphosis of the ephyrae is more complicated than in the latter No species of this group is injurious to the human skin Some large forms such as *Rhopilema are used as food in* China *and Japan*

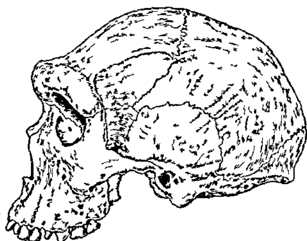
A fair number of fossils of this order were found in the strata of the Jurassic Period *See* SCYPHOZOA [T U]



Rhizostomeae (a) *Rhizostoma* bell may reach 2 ft in diameter (b) *Mastigias* (L. Hyman *The Invertebrates* vol. 1 McGraw Hill 1940)

Rhodesian man

A type of fossil man inhabiting southern Africa during the late Pleistocene presumably more than 40 000 years ago The first specimens were dis-covered in mining operations at Broken Hill North ern Rhodesia in 1921 They consisted of a skull without lower jaw a second upper jaw fragment *and bones of the limbs and pelves from* more than one individual These lay at the end of a deep filled cave associated with a relatively recent fauna (containing few extinct species) and tools of the *Rhodesian Proto Still Bay culture* of the Early Middle Stone Age of Africa



Skull of Rhodesian man *(From M F Ashley Montagu An Introduction to Physical Anthropology 2d ed Charles C Thomas 1951)*

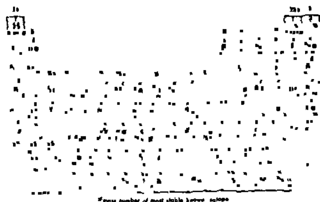
A second skull cap in fragments was found by R Singer in 1953 It was located at a blowout site at *Hopefield* near Saldanha Bay Cape Colony This was associated with an older fauna and stone work of the Earlier Stone Age however no great difference in time between these and the Broken Hill specimens is indicated

The skulls are large and low, marked by massive brow ridges, and have a cranial capacity of approximately 1300 cm³ or less. In detail, Rhodesian man differs from the contemporary Neanderthals in conformation of the skull vault and the brows, which are not arched as in the Neanderthals, and of the palate, which, although large, is parabolic in shape rather than U shaped. The skeleton bones differ in no particular from modern man. The type has been named *Homo rhodesiensis* (Woodward) and *Cyphanthropus rhodesiensis* (Pycraft). *See* FOSSIL MAN [W W H]

Bibliography W P Pycraft, G E Smith, M Yearsley et al, *Rhodesia Man and Associated Remains*, 1928

Rhodium

Chemical element number 45, rhodium, Rh, is a hard white metal. It is considerably less ductile than either platinum or palladium but much more ductile than any of the other platinum metals. Of the platinum metals, rhodium and platinum are the most suitable for use in hot, oxidizing atmospheres.



Uses. *Pure rhodium can be electroplated easily* to form a hard wear resistant permanently bright surface. The plated metal is used on sliding electrical contacts for mirrors and reflectors and as a finish on jewelry. It is also employed in furnace windings and crucibles for use at temperatures too high for platinum. Rhodium catalysts have been used in certain organic syntheses.

Alloys. The major rhodium alloy is 90% platinum 10% rhodium. The addition of rhodium improves the hardness, strength, and corrosion resistance of platinum. The alloy is used for thermocouples, furnace windings, parts of high temperature apparatus, jets in making synthetic fibers, and as a catalyst in the oxidation of ammonia to produce nitric acid. Rhodium platinum alloys are preferred to other platinum metal alloys for most high temperature oxidizing exposures and are essential for the production of fiber glass.

Metallurgical extraction. In one method of extraction, the major portions of platinum and palladium are first removed chemically from the other platinum metals, and the residue is dissolved in molten lead. The lead dissolves platinum, palladium, and rhodium, and the platinum and palladium are removed by dissolving the alloy in nitric acid. Heating the impure rhodium in chlorine causes the impurities to sublime, insoluble rhodium chloride is left. The rhodium is extracted from this residue by a sodium bisulfate fusion followed by leaching with water. Rhodium chloride is then precipitated from the resulting solution, and after further treatment, rhodium metal is obtained by reduction with hydrogen.

Properties of rhodium

Atomic weight	102 91
Crystal structure	face-centered cubic $a = 3.80$ at 20°C
Density (at 20°C)	12 41 g/cm³
Melting point	1960°C
Linear thermal expansion coefficient (per °C at 20°C)	8.3×10^{-6}
Specific heat (at 0°C)	0 059 cal/°C
Thermal conductivity (at 20°C) (cal)/(cm)(cm²)(°C)(sec)	0 21
Electrical resistivity (at 20°C)	4 5 μohm-cm
Modulus of elasticity	40×10^6 psi
Tensile strength—hard	365 000 psi
Tensile strength—annealed	138 000 psi

Physical and chemical properties. When hot rhodium is quite ductile, when cold, the metal still has appreciable ductility although it work hardens very rapidly. It can be made into fine wire and thin sheet metal.

Rhodium is the whitest of the platinum metals, it remains bright at ordinary temperatures under all atmospheric conditions. Rhodium oxidizes slightly when heated but above 1000°C the oxide, Rh_2O_3 decomposes. Rhodium is resistant to most common acids including hot aqua regia, even at moderate temperatures. It is attacked by hot sulfuric acid hot hydrobromic acid, sodium hypochlorite and free halogens at 200–600°C.

Rhodium compounds. Rhodium exhibits oxidation states of 3+ and 4+, the former being the principal one. Rhodium trichloride, $RhCl_3$, is a red compound which is insoluble in water. Its low volatility makes it useful in the refining of the element. Rhodium trihydroxide may be formed from the trichloride by boiling with potassium hydroxide. This is soluble in some acids and may be used to produce rhodium salts. Rhodium sulfate, $Rh_2(SO_4)_3 \cdot xH_2O$, is red or yellow and soluble in water. It appears to become complex after being subjected to high temperatures, as such it forms the basis of rhodium plating baths. *See* PLATINUM [H J A]

Rhodobacteriineae

A suborder of the order Pseudomonadales which contains all photosynthetic bacteria except the genus *Rhodomicrobium*. The unusual developmental history of this genus has caused it to be included in the order Hyphomicrobiales. *See* HYPHOMICROBIALES, PSEUDOMONADALES

Photosynthesis in the Rhodobacteriineae differs from that of green plants in the following respects (1) it never leads to oxygen evolution (2) it re-

ceeds in the near infrared region 700–1000 millimicrons (mμ) The radiant energy is absorbed by chlorophyllous pigments other than those found in the green plants. *See* CHLOROPHYLL, PHOTOSYNTHESIS

The Rhodobacteriineae comprise three families the Thiorhodaceae, Athiorhodaceae, and Chlorobacteriaceae, whose representative genera are shown in the illustration

Thiorhodaceae These are the purple sulfur bacteria. They are often found as mass developments in mud or water containing H_2S They are anaerobic, that is cannot grow in the presence of oxygen, and contain bacteriochlorophyll and purple, red, and yellow carotenoids The bacteriochlorophyll has its major absorption maximum at about 900 mμ the carotenoids mask its greenish blue color so that the bacteria appear purplish to red colored.

As long as H_2S is present in the environment, the cells contain sulfur globules the result of an in

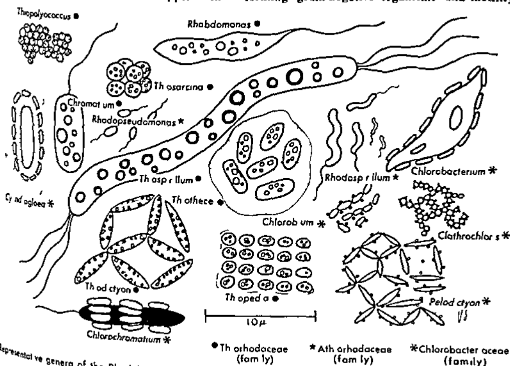
when sulfur free they are deep red

A simple and effective method for cultivating Thiorhodaceae in the laboratory is the following. Some black mud, or a bit of material from a natural mass development of these bacteria, is put into a glass-stoppered bottle which is then filled completely with a mineral medium The stopper is inserted carefully so that no air bubbles are left, and the bottle is exposed to continuous illumination by a 25- to 40 watt electric bulb Fluorescent lights, which emit very little infrared radiation, are unsatisfactory Experience has shown that good cultures can usually be obtained in 3-6 days with a medium composed of tap water with 0 1% ammonium chloride (NH_4Cl), 0.1% monobasic potassium phosphate (KH_2PO_4), 0 05% magnesium chloride ($MgCl_2$), 0 2% sodium bicarbonate ($NaHCO_3$), and 0 05% sodium sulfide ($Na_2S \cdot 9H_2O$) adjusted to pH 7–9 For brackish and salt water forms the addition of 1–2% sodium chloride (NaCl) may be necessary Owing to the toxicity of the sulfide media with more than 0 1% sodium sulfide (Na_2S) may fail to yield growth of purple sulfur bacteria Substitution of 0 1–0 2% sodium thiosulfate for Na_2S or supplementation of the sulfide medium with thiosulfate is apt to produce denser and more varied cultures. The sulfide concentration and the pH of the medium often determine which types of Thiorhodaceae will appear in the cultures

Although H_2S is the preferred oxidizable substrate for the purple sulfur bacteria they can also use a variety of simple organic compounds such as ethanol, salts of fatty acids lactate, malate and so forth

Many representatives of the group are motile, these display photoactic movements whereby the bacteria accumulate in areas of optimum light intensity for their photosynthetic activity *See* BACTERIAL MOTILITY TAXIS

The Thiorhodaceae comprise various types of spherical ovoid comma rod and corkscrew-shaped bacteria of diverse sizes They are nonsporeforming gram negative organisms and motility is



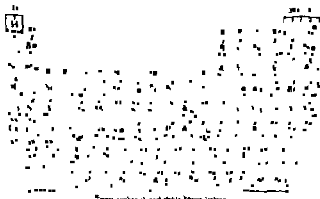
Representative genera of the Rhodobacteriineae (V B D Skerman)

The skulls are large and low, marked by massive brow ridges, and have a cranial capacity of approximately 1300 cm³ or less In detail, Rhodesian man differs from the contemporary Neanderthals in conformation of the skull vault and the brows, which are not arched as in the Neanderthals, and of the pa' ' '
shape ra ' . .
differ in .
has been . :
and *Cyphanthropus rhodesiensis* (Pycraft) *See* FOSSIL MAN [W W H]

Bibliography W P Pycraft, G E Smith, M Yearsley et al, *Rhodesia Man and Associated Remains*, 1928

Rhodium

Chemical element number 45, rhodium, Rh, is a hard white metal It is considerably less ductile than either platinum or palladium but much more ductile than any of the other platinum metals Of the platinum metals, rhodium and platinum are the most suitable for use in hot, oxidizing atmospheres



Uses. Pure rhodium can be electroplated easily to form a hard, wear resistant, permanently bright surface The plated metal is used on sliding electrical contacts, for mirrors and reflectors, and as a finish on jewelry It is also employed in furnace windings and crucibles for use at temperatures too high for platinum Rhodium catalysts have been used in certain organic syntheses

Alloys. The major rhodium alloy is 90% platinum 10% rhodium The addition of rhodium improves the hardness strength, and corrosion resistance of platinum The alloy is used for thermocouples, furnace windings, parts of high temperature apparatus, jets in making synthetic fibers and as a catalyst in the oxidation of ammonia to produce nitric acid Rhodium platinum alloys are preferred to other platinum metal alloys for most high temperature oxidizing exposures and are essential for the production of fiber glass

Metallurgical extraction. In one method of extraction, the major portions of platinum and palladium are first removed chemically from the other platinum metals, and the residue is dissolved in molten lead The lead dissolves platinum, palladium, and rhodium, and the platinum and palladium are removed by dissolving the alloy in nitric acid Heating the impure rhodium in chlorine causes the impurities to sublime, insoluble rhodium chloride is left The rhodium is extracted from this residue by a sodium bisulfate fusion followed by leaching with water Rhodium chloride is then precipitated from the resulting solution and after further treatment, rhodium metal is obtained by reduction with hydrogen

Properties of rhodium

Property	Value
Atomic weight	102 91
Crystal structure	face-centered cubic a = 3 80 at 20°C
Density (at 20°C)	12 41 g/cm³
Melting point	1960°C
Linear thermal expansion coefficient (per °C at 20°C)	8.3×10^{-6}
Specific heat (at 0°C)	0 059 cal/°C
Thermal conductivity (at 20°C) (cal)/(cm)(cm²)(°C)(sec)	0 21
Electrical resistivity (at 20°C)	4 5 µohm-cm
Modulus of elasticity	40×10^6 psi
Tensile strength—hard	365 000 psi
Tensile strength—annealed	138 000 psi

Physical and chemical properties. When hot, rhodium is quite ductile, when cold, the metal still has appreciable ductility although it work hardens very rapidly It can be made into fine wire and thin sheet metal

Rhodium is the whitest of the platinum metals, it remains bright at ordinary temperatures under all atmospheric conditions Rhodium oxidizes slightly when heated but above 1000°C the oxide, Rh_2O_3, decomposes Rhodium is resistant to most common acids including hot aqua regia, even at moderate temperatures It is attacked by hot sulfuric acid hot hydrobromic acid, sodium hypochlorite, and free halogens at 200–600°C

Rhodium compounds. Rhodium exhibits oxidation states of 3+ and 4+, the former being the principal one Rhodium trichloride, $RhCl_3$, is a red compound which is insoluble in water Its low volatility makes it useful in the refining of the element Rhodium trihydroxide may be formed from the trichloride by boiling with potassium hydroxide This is soluble in some acids and may be used to produce rhodium salts Rhodium sulfate, $Rh_2(SO_4)_3 \cdot xH_2O$, is red or yellow and soluble in water It appears to become complex after being subjected to high temperatures, as such it forms the basis of rhodium plating baths *See* PLATINUM [H J A]

Rhodobacteriineae

A suborder of the order Pseudomonadales which contains all photosynthetic bacteria except the genus *Rhodomicrobium* The unusual developmental history of this genus has caused it to be included in the order Hyphomicrobiales *See* HYPHOMICROBIALES PSEUDOMONADALES

Photosynthesis in the Rhodobacteriineae differs from that of green plants in the following respects (1) it never leads to oxygen evolution, (2) it re

The yellow green color is *due to the presence* of yet another kind of chlorophyll, chlorobium chlorophyll, whose major absorption maximum lies at 30–40 mμ and to the accompanying yellow carotenoids. Most representatives are nonmotile, but motile types have been observed though not yet

wise may easily overgrow the green sulfur ...
The latter can also be selectively cultivated by making use of the specific absorption maximum of the chlorobium chlorophyll. For this purpose the bottle cultures are exposed to radiation in the region of 00–60 mμ here the purple bacteria cannot grow. This radiation can be simply achieved by using light from incandescent bulbs passed through filters that absorb the radiation at wavelengths below 00 and above 60 mμ.

Various types of green sulfur bacteria have been found in natural collections, they range from small spherical or ovoid to distinctly rod shaped cells. Subdivision into genera is as yet somewhat arbi

environment. As in the case of the Thiorhodaceae much more work with pure cultures is needed in order to develop the principles on which a more satisfactory system of classification can be based. The genera recognized at present are described in the following paragraphs.

Chlorobium is composed of small spherical cells ...

or less irregular long rods or as tightly wound coils.

Pelodictyon species have ovoid to distinctly rod shaped cells united into large colonies of charac

in parallel strands or as irregularly arranged cell masses.

Clathrochloris is similar to *Pelodictyon* and is composed of spherical cells in chains which in turn ...

... the cells.

Microchloris occurs as single, thin, rod shaped cells.

Two further genera *Chlorochromatium* and *Cylindrogloea* represent combinations of green sulfur bacteria with presumably, colorless bacteria. *Chlorochromatium* also known as *Chloronium mirabile* grows *in* the form of barrel shaped structures, composed of a rod shaped motile central bacterium *covered entirely* by a single layer of ovoid green sulfur bacteria. The combination divides *synchronously so that the typical shape of the composite is perpetuated. The probably colorless central bacterium confers motility on the complex.*

Cylindrogloea is composed of large, cylindrical aggregates of a filamentous central bacterium, surrounded by a layer of ovoid to rod shaped green sulfur bacteria.

Chlorobacterium is the name given to green presumably sulfur bacteria that grow as rod shaped, often slightly curved cells forming an outside covering on cells of protozoa notably of amebas and flagellates. *See* BACTERIA TAXONOMY OF

[C.B.V.N.]

Bibliography W Bavendamm *Die Schwefelbakterien*, 1924. H Larsen *On the Microbiology and Biochemistry of the Photosynthetic Green Sulfur Bacteria* 1952. E G Pringsheim *Archiv f Mikrobiol.*, 19 353 1953. C B van Niel *Archiv f Mikrobiol* 3 1 1931. C B van Niel *Bacteriol Revs.*, 8 1, 1944. S Winogradsky *Microbiologie du sol*, 1949.

Rhodochrosite

The mineral form of manganese carbonate. Calcium iron magnesium and zinc have all been reported to replace some of the manganese. The equilibrium replacement of manganese by calcium has been determined and found to increase with the temperature of crystallization.

Rhodochrosite is sometimes found in low temperature veins near deposits of copper lead zinc and silver or it may occur with other manganiferous minerals of higher temperature origin. It has also been found in sediments and in pegmatites. Well known occurrences of rhodochrosite are in Europe Asia and South America. In the United States large quantities occur at Butte Montana. As a source of manganese rhodochrosite is also important at Chamberlain South Dakota and in Aroostook County Maine.

Rhodochrosite has hexagonal (rhombohedral) symmetry and the calcite type crystal structure. It occurs more often in massive or columnar form than in distinct crystals. The color ranges from pale pink to brownish pink. Hardness is 3½–4 on Mohs scale and specific gravity is 3 70. Rhodochrosite is stable up to 690°C at 10 000 psi and 790°C at 29 000 psi of carbon dioxide. It can be synthesized in its stability field. *See* CARBONATE MINERALS; MANGANESE.

[R.I.H.]

Bibliography J R Goldsmith and D L Graf The system CaO MnO CO_2 solid solution and decomposition relations *Geochim et Cosmochim Acta* 11 310–334 1957

Rhodonite

... SiO_3 ... n in ... the base. More often it is in cleavable to compact masses or in imbedded grains. Crystallographically

due to the presence of polar flagella When found in nature the cells are often grouped in characteristic aggregates which may be embedded in a common slime envelope (capsule) The fundamental studies of S Winogradsky who first demonstrated the importance of H_2S for the growth of the sulfur bacteria led him to subdivide the group into a number of genera differentiated on the basis of the shape and size of the individual cells and of the aggregates Subsequent studies have shown however that pure cultures of a single strain may contain cells and aggregates of many different shapes and sizes these have been correlated with the gradually changing conditions in the culture medium Althoug ty of many o clas sificatio e satisfactory basis for a subdivision can be expected as the result of the isolation in pure culture of the different types encountered in nature, followed by extensive comparative studies on the behavior of such pure cultures under a wide variety of environmental conditions Thus far only a few types have been so examined The genera of the Thiorhodaceae that are recognized at present are described in the following paragraphs

Thiosarcina is composed of spherical cells combined into regular cubical packets of 8–64 cells They result from consecutive cell divisions in three perpendicular planes Smaller units in the form of squares of four cells also known as tetrads may also be found *Thiosarcina* is sometimes motile

Thiopedia forms flat sheets of nonmotile spherical cells The sheets are of various dimensions and may resemble those of the blue green alga *Merismopedia* and of the colorless bacterium *Lampropedia* Under conditions of rapid growth the organisms often appear as single individuals or as groups of 2–4 cells in that event a distinction from the non motile essentially unicellular *Rhodothece* becomes difficult if not impossible

Thiopolycoccus characteristically form dense irregular aggregates of nonmotile spherical cells not embedded in a common capsule

Thiocystis is composed of highly motile slightly ovoid cells often in pairs and forming small rather dense aggregates enclosed in a common capsule The very similar *Thiothece* is differentiated because the aggregates are much more sparsely populated

Lamprocystis also resembles *Thiocystis* rather closely it is distinguished because during growth the aggregates break up into small clusters which form a reticulum

Thiodictyon is a genus in which the rod shaped cells are arranged end to end producing netlike aggregates Groups of cells may separate from the main colony and move away

Amoebobacter species are composed of spherical to ovoid cells combined into large aggregates of irregular structure with characteristic hollow spaces in them The colonies display changes in shape owing to the motility of the component cells even though the cell masses have a tendency to remain united

Chromatium is usually found as single rod to comma shaped motile cells It is perhaps the most common of the Thiorhodaceae and contains the largest representatives among the sulfur purple bacteria with cells measuring up to 6–10 by 15 20 microns (μ)

Rhabdomonas also known as *Rhabdochromatium* occurs as irregularly shaped cells usually elongated with tapering ends and swollen distorted central portions Such structures have also been found in pure cultures of *Chromatium* species growing under unfavorable conditions hence *Rhabdomonas* may represent an aberrant form of *Chromatium*

Thiospirillum comprises the typical corkscrew shaped members of the purple sulfur bacteria Among them are some of the largest spirilla known such as *Thiospirillum jenense* and *T sanguinea* which are sometimes found as mass developments in natural environments and measure 3–4 by 30–40 μ or even 100 μ None of the *Thiospirillum* species has yet been cultivated in the laboratory

Athiorhodaceae These are the nonsulfur purple and red bacteria They do not oxidize sulfide sulfur or thiosulfate and are dependent for growth on organic substances Some representatives can use molecular hydrogen as the oxidizable substrate for photosynthesis but even these cannot develop in strictly inorganic media because all types thus far studied require one or more B vitamins for growth (*see* VITAMIN) The organisms contain the same bacteriochlorophyll as the Thiorhodaceae accompanied by various red purple and yellow carotenoids the latter determining the color of the cultures

The group comprises both strictly anaerobic and facultatively anaerobic members The former can grow only when exposed to light the latter can also develop in darkness if the cultures are exposed to air

Laboratory cultures of the Athiorhodaceae can readily be obtained by a slight modification of the medium specified for the cultivation of Thiorhodaceae In place of sulfide the medium is supplemented with 0 1–0 2% of an organic substance such as ethanol salts of fatty acids lactate and malate and with a complement of B vitamins or with a small amount of yeast extract which serves as a source of these vitamins

The Athiorhodaceae represent a group of non sporeforming gram negative polarly flagellated small rods and spirilla The rod shaped representatives are classified as *Rhodopseudomonas* species the spirilla as *Rhodospirillum* species

Chlorobacteriaceae These are the green sulfur bacteria They have been found in nature only in sulfide containing environments and are frequently associated with Thiorhodaceae They are strict anaerobes and the few types that have been isolated and studied in pure culture can use sulfide sulfur and molecular hydrogen as oxidizable substrates Only one representative is known to grow with organic substances instead of sulfide in the medium

Reproduction. The reproductive structures of this group are so different from the other algae that a special series of terms is used to describe them. There are no flagellated reproductive cells. The female reproductive organ, corresponding to the oogonium in other algae, consists of a swollen basal portion known as a carpogonium, containing ... slender projection, the ...

... nonmotile male cell called a spermatium. Since the naked male cells, or spermatia, have no motility of their own, they are carried by the movements of the water. When, by chance, a spermatium becomes

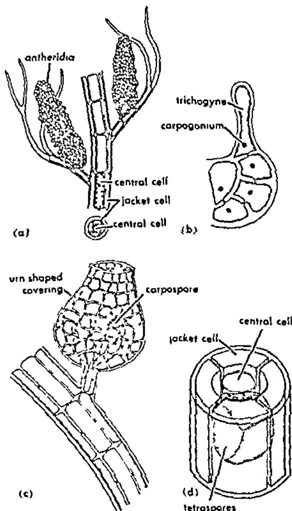


Fig 3. *Polysiphonia* (a) Branched thallus with single row of central cells surrounded by jacket cells. Two clusters of antheridia are shown on branches. Cross section of axis showing single central cell surrounded by jacket cells. (b) Carpogonium with elongated trichogyne. Note prominent cytoplasmic connections between cells. (c) Carpospores surrounded by urn-shaped covering. (d) Portion of axis showing tetraspores lying between row of central cells and jacket cells. (From H. J. Fuller and O. Tippo, *College Botany*, rev. ed., Holt, 1954)

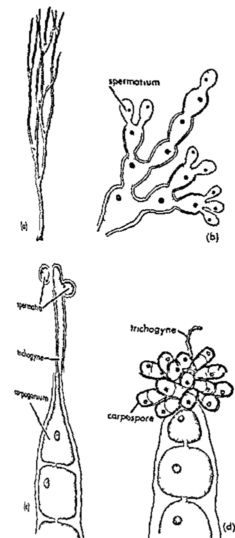


Fig. 2. *Nemalion* (a) Thallus is cylindrical and forked. (b) Brushlike filaments bearing terminal spermatangia, each of which contains one spermatium. (c) Filament with a terminal carpogonium (containing an egg) and bearing an elongated trichogyne. Two spermatia are shown at the tip of the latter. (d) With carpospores. (From H. J. Fuller and O. Tippo, *College Botany*, rev. ed., Holt, 1954)

attached to a ... ogyne walls br... enters the trichogyne, ... the carpogonium where it fuses with the carpogonial nucleus. In *Nemalion* the resulting zygote is retained by the female plant and it soon undergoes nuclear as well as cell divisions forming asexual spores known as carpospores (Fig. 2). In *Polysiphonia* and *Griffithsia* the carpospores are formed in a rather complicated postfertilization development (Fig. 3). In these species each carpospore germinates into a diploid plant, the tetrasporophyte, which is similar in general appearance and size to the male and female gametophytes but at maturity produces a tetrasporangium. In two suc

rhodonite is closely related to the pyroxenes and thus has two cleavage directions at about 88 and 92° Hardness is 5 5–6 on Mohs scale and specific gravity is 3 4–3 7 The luster is vitreous and the color is rose red pink or brown Rhodonite is similar in color to rhodochrosite manganese carbonate but it may be distinguished by its greater hardness and insolubility in hydrochloric acid It has been found at Langban Sweden near Sverdlovsk in the Ural Mountains and at Broken Hill Australia Fine crystals of a zinc bearing variety fowlerite are found at Franklin New Jersey *See* SILICATE MINERALS [C S HU]

Rhodophyta

A large phylum of small to medium sized plants which are also called red algae The phylum contains 2500 species all of which are marine except about 50 that grow in fresh water They are most abundant in the tropics where in clear water they *are found growing at a depth of 600 ft* The majority of the species grow on rocks or other solid substrata, although there are a number of species which grow on other algae as epiphytes endophytes or partial parasites There is a wide variation in the structure and appearance of the plants of this group some being beautiful in either the fresh or the dried state

Economic importance The red algae have a wide range of uses *Porphyra* highly esteemed in the Orient is widely cultivated in some of the quieter estuaries and tidelands of Japan where the sun dried product is used both as a condiment and in soups Irish moss or carrageen (*Chondrus crispus*) is harvested to a certain extent along the coast of the northern Atlantic the gel (car [illegible]

[illegible] (*Rhody*

menia palmata) has limited use as a food and relish However the most important contribution of members of this phylum is the production of agar which is obtained from various species of *Gelidium* and other red algae Agar is used extensively in medicine as a bulk producing laxative and as a medium for the culture of bacteria and fungi The red algae also include coralline or calcareous forms which contribute appreciably to the building up of coral reefs

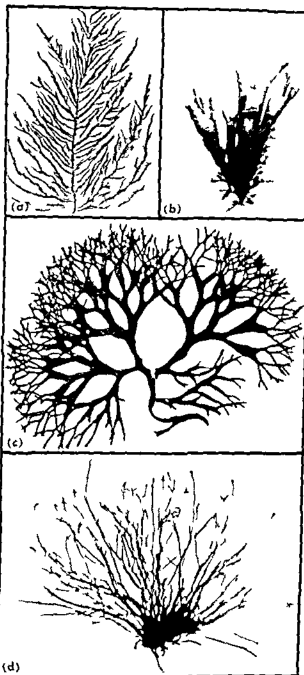


Fig 1 (a) *Dasya* red alga with delicate thallus (b) *Rhodymenia* or dulse with a sheetlike blade (c) *Chondrus crispus* or Irish moss (d) *Gelidium* one of the red algae from which agar is obtained (From H J Fuller and O Tippo *College Botany* rev ed Holt 1954)

Pigmentation Phycoerythrin is usually present in such abundance that it masks the chlorophylls and carotenoids and gives the plants a range of shades *of red which accounts for their common* name red algae Sometimes phycocyanin is also present giving the plants a bluish to almost black color The variation in the amounts of these different pigments accounts for the wide diversity of colors to be found in the group There is a close correlation between the habitat and color The deeper occurring marine species that are never exposed by the tides have the most characteristic pink or bright red color whereas those growing in the upper tidal area or in fresh water rarely have the characteristic red color but are usually various shades of olive brown dull green or bluish to black The red algae *may be found at greater oceanic* depths than other algae They seem to be able to utilize the deeper penetrating blue rays of the spectrum for photosynthesis more effectively than other plants The pigments are confined to plastids which may be large and single in each cell or small and numerous The excess photosynthate is stored in the form of a carbohydrate of the general nature of glycogen known as floridean starch which gives a wine-red stain when treated with iodine potassium iodide solutions

Rhynchocephalia

An order of lizardlike reptiles represented today by a single living species the tuatara, *Sphenodon punctatus*, of the Cook Strait and Bay of Plenty Islands just off New Zealand The group first appeared in the fossil record in the Triassic and flourished through the Jurassic but no fossils are known subsequent to that time The living form is a moderately large reptile attaining a length of 2½ ft and a weight of 2 lb It is readily distinguished from other living reptiles by having the following combination of characters, two temporal fenestrae, an immovable quadrate, no secondary palate, no shell, and no penis in the males In most characteristics *Sphenodon* resembles rather closely the primitive fossil members of the order, which are placed with it in the family Sphenodontidae The tuatara is a secretive animal active at body temperatures of 43–56°F, and spends its days in shallow burrows At night it emerges to feed on terrestrial snails and insects and in season, to take part in courtship activities Correlated with the

eggs which are buried in the burrow Incubation takes 12 13 months The young expedite hatching by cutting their way out of the leathery egg with a small egg tooth, actually a horny caruncle, that develops at the tip of the snout and is lost shortly after they escape *See* REPTILIA, *See also* LEPIDOSAURIA

[J M S]

Rhynchocoela

A phylum of bilaterally symmetrical unsegmented, ribbonlike worms frequently referred to as the nemertinea They have an eversible proboscis and a complete digestive tract with an anus There is no coelom or body cavity, and the mesenchyme or parenchyma and the muscle fibers fill the area between the ciliated epidermis and the cellular lining of the digestive tract *See* ACOELOMATA, ANIMAL SYMMETRY

Morphology. The nemertineans are mostly less than 20 cm in length but a few may reach a length of several meters Many species are brightly colored, sometimes having stripes or transverse bars The tubular proboscis, lying above the digestive tract in a cavity the rhynchocoele, is attached posteriorly to the proboscis sheath by a retractor muscle and is either unarmed or armed with stylets The proboscis opens anteriorly into a chamber the rhynchodeum, which in turn opens to the outside above the mouth, through the proboscis pore The proboscis can be suddenly everted by the contraction of muscles which exert pressure on the fluid in the rhynchocoele It is used for capturing prey, such as annelid worms, and for defense for locomotion, or for burrowing

The nemertineans constitute the most primitive group of in ... in which the digestive tract ... and anus In some nemer

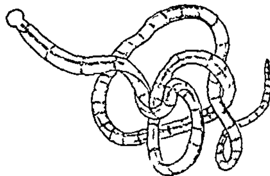


Fig 1 *Tubulanus capistratus from the Pacific coast; size of various species ranges from 5 to 30 mm (From L H Hyman, The Invertebrates, vol 2, McGraw Hill 1951)*

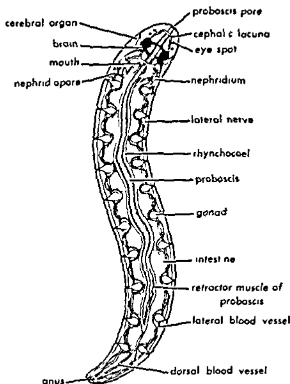


Fig 2 Diagram of the internal structure of a nemertinean

tineans however, a separate mouth is lacking and the esophagus opens through the proboscis pore

The nemertineans are the simplest animals with a circulatory system There are two lateral blood vessels and in some a third unpaired dorsal vessel *The blood consists of a colorless fluid which may* contain blood cells of several types In species where the blood is colored the pigment is present in the cells There is no heart but the walls of the principal vessels may be contractile

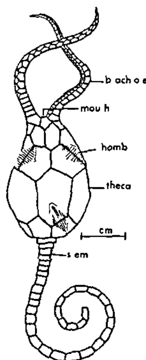
The excretory organs or protonephridia are composed of many tubules ending in flame bulbs These are hollow urn shaped cells provided with vibratile cilia On each side the flame bulbs are often closely associated with the lateral blood vessel and their tubules are united to a common tubule which opens at a lateral nephridiopore *Peculiar* large cells

cessive nuclear divisions meiosis is accomplished so that the resulting tetraspores are haploid (*see* MEIOSIS) These have been observed to develop into gametophytic plants Thus involved in the life cycles are three different somata (bodies) the haploid gametophytes being either male or female the carposporophyte or diploid stage and the free living diploid stage or tetrasporophyte which reproduces by tetraspores from which the haploid gametophytes arise This type of life cycle is found in no other group of plants except certain ascomycetous fungi (*see* FUNGI) The phylum appears to be a specialized terminal series in evolution which leads to no other group [P A V]

Bibliography *See* THALLOPHYTA

Rhombifera

An order of Cystidea in which the thecal canals crossed the sutures at the edges of the plates so that one half of any canal lay in one plate and the



Pleurocyst s from the Ordov c an (S mpl fed after O Jaekel)

other half on an adjoining plate The canals sometimes occurred in lozenge shaped clusters called rhombs The theca was ovoid and sessile in earlier forms and comprised numerous irregular plates The ambulacral grooves were trimerous and were restricted to the brachioles In later forms the thecal plates became fewer and larger and were arranged in five horizontal rows termed cycles A stem developed aborally and the ambulacra became pentamerous traversing part of the theca before ascending the brachioles Rhombifera probably died out in the Devonian *See* CYSTIDEA [H B F]

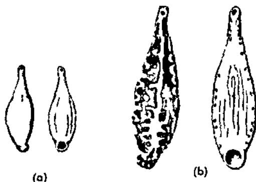
Rhubarb

An herbaceous perennial (*Rheum rhaponticum*) of Mediterranean origin belonging to the plant order Polygonales Rhubarb is grown for its thick petioles which are used mainly as a cooked dessert Propagation is by division of root crowns Victoria Macdonald and Valentine are popular varieties Commercial production is generally limited to areas where crowns may become dormant for 2 3 months each year Outdoor rhubarb is a common garden vegetable in most areas of the United States except the South Harvesting begins in the spring and continues for 6–10 weeks Commercial plantings are renewed every 4–8 years Michigan and Washington are important centers for forced or hothouse rhubarb Two or three-year old field grown crowns are moved into darkened forcing structures in late winter and forced at 55–60°F to obtain petioles of a bright red color *See* POLYGONALES VEGETABLE GROWING [H J C.]

Rhynchobdellae

An order of the class Hirudinea These leeches possess an eversible proboscis and lack hemoglobin in the blood They may be divided into two families the Glossiphoniidae and the Ichthyobdellidae

Glossiphoniidae are flattened mostly small leeches occurring chiefly in fresh water *Theromyzon* is a parasite of water fowl sucking blood from the nasal cavity the mouth or the eye A heavy infestation on a young bird may be fatal *Hemiclepsis* sucks the blood of fishes and amphibians and is found in stagnant waters in Europe and northern Asia *Placobdella* is a common American parasite of aquatic turtles frogs and fishes and *Glossiphonia* sucks the blood of acquatic invertebrates especially snails The genus *Haementeria* contains reptile parasites chiefly but also includes parasites of mammals including man *Haementeria officinalis* is used in Mexico for medicinal purposes



Rhynchobdellae (a) Gloss phon a complanata (b) Placobdella paras t ca

Ichthyobdellidae typically have cylindrical bodies with conspicuous powerful suckers used to attach themselves to passing fish They frequently have lateral appendages which aid in respiration There is a tendency for a particular species to be confined to a specific host For instance *Callobdella lophii* is confined to the angler fish *Lophius piscatorius* and *Abranchus microstomus* to the small shore fish *Cottus scorpius* On the other hand *Pontobdella muricata* attacks a variety of skates and rays *Crangonobdella* is an example of a parasite of crustaceans *Piscicola* is one of the few fresh water forms *See* ARHYNCHOBDELLAE HIRUDINEA [K H M]

Rhynchocephalia

An order of lizardlike reptiles represented today by a single living species the tuatara *Sphenodon punctatus* of the Cook Strait and Bay of Plenty Islands just off New Zealand The group first appeared in the fossil record in the Triassic and flour ... the Jurassic but no fossils are ... is ... of ... in ... guished from other living reptiles by having the following combination of characters, two temporal fenestrae an immovable quadrate no secondary palate no shell and no penis in the males In most cha ... the ... are placed with it in the family ...

The tuatara is a secretive animal active at body ...

part in courtship activities Correlated with the nocturnal habits are a vertically elliptical pupil and also a small croaking voice probably used in territorial or breeding activity Fertilization is internal and the female produces a number of small eggs which are buried in the burrow Incubation takes 12 13 months The young expedite hatching by cutting their way out of the leathery egg with a small egg tooth actually a horny caruncle that develops at the tip of the snout and is lost shortly after they escape See REPTILIA See also LEPIDOSAURIA. [J M S]

Rhynchocoela

A phylum of bilaterally symmetrical unsegmented ribbonlike worms, frequently referred to as the ...

of the digestive tract See ACOELOMATA ANIMAL ...

Morphology The nemertineans are mostly less than 20 cm in length but a few may reach a length of several meters. Many species are brightly colored sometimes having stripes or transverse bars The tubular proboscis lying above the digestive tract in a cavity the rhynchocoele is attached posteriorly to the proboscis sheath by a retractor muscle and is either unarmed or armed with stylets The proboscis opens anteriorly into a chamber the rhynchodeum which in turn opens to the outside ... The ... trac ... fluid ... rhynchocoele. It is used for capturing prey mostly annelid worms and for defense for locomotion or for burrowing.

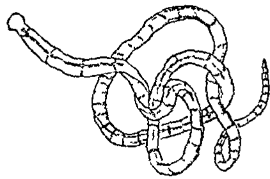


Fig 1 *Tubulanus capistratus* from the Pacific coast size of various species ranges from 5 to 30 mm (From L H Hyman, The Invertebrates, vol 2 McGraw Hill 1951)

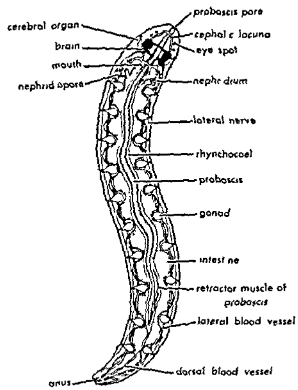


Fig 2 Diagram of the internal structure of a nemertinean

tineans however a separate mouth is lacking and the esophagus opens through the proboscis pore

The nemertineans are the simplest animals with a circulatory system There are two lateral blood vessels and in some a third unpaired dorsal vessel. The blood consists of a colorless fluid which may contain blood cells of several types In species where the blood is colored the pigment is present in the cells There is no heart but the walls of the principal vessels may be contractile

The excretory organs or protonephridia are composed of many tubules ending in flame bulbs These are hollow urn shaped cells provided with vibratile cilia On each side the flame bulbs are often closely associated with the lateral blood vessel and their tubules are united to a common tubule which opens at a lateral nephridiopore Peculiar large cells

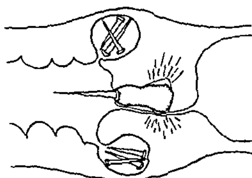


Fig 3 Stylets of *Amphiporus griseus* (From W R Coe, *Biology of the nemerteans of the Atlantic coast of North America, Trans Conn Acad Arts Sci*, 35 282, 1943)

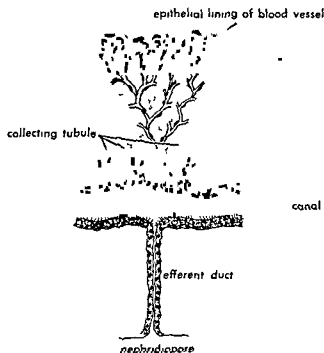


Fig 4 Diagram of the multiple protonephridium of a nemertinean (From W R Coe *Unusual types of nephridia in nemerteans, Biol Bull*, 58(3) 214, 1930)

called athrocytes may surround the flame bulbs and tubules and are assumed to be excretory since they readily take up vital stains

The nervous system has a pair of cerebral ganglia forming the brain as well as two longitudinal nerve cords and many smaller nerves The ganglia and lateral cords may contain unusually large neurochord cells In the epidermis there are scattered sensory nerve cells, probably tactile Chemotactile organs are situated in a pair of anteriorly placed cephalic grooves or in a flask shaped protrusible frontal organ A few to many simple eyes or ocelli may be present in front of the cerebral ganglia Statocysts, the organs of equilibrium are rare Two blind canals, the cerebral organs, are invaginated from the epidermis and are closely associated with the cerebral ganglia These organs probably chemosensory, open through pores in the cephalic grooves or on the body surface

There are no special respiratory organs and respiration occurs through the body surface

Nemertineans are usually either male or female but a few individuals have both sex organs The ovaries or testes open by short ducts to the exterior Fertilization occurs outside the body in many species but may be internal in certain forms

Embryology. Cleavage of the fertilized egg or zygote, is spiral, with the cells of the lower quartet of the 8 cell stage rotated slightly, to lie in the furrows between the cells of the upper quartet Development is determinate, with the potentialities for future development of the embryo determined or fixed in the zygote before cleavage begins Isolated cells of the 2 cell stage result in dwarf larvae Isolated cells of later stages result in deficiencies *See* CELL LINEAGE, CLEAVAGE, EMBRYONIC

After early cleavage, the development in certain nemertineans may be direct, that is, without a larva the embryo emerging from the egg membranes as a minute, ciliated worm In others, the gastrula becomes a free swimming, helmet shaped, ciliated pilidium, formed by the downward growth of two ciliated lobes at the sides of the mouth and having an apical tuft of cilia In still other nemertineans the gastrula remains inside the egg membranes and becomes an oval, ciliated larva known as Desor's larva, which lacks the apical tuft and the oral lobes

Both the pilidium and the Desor's larva metamorphose into an adult worm by means of the invagination of seven or eight ectodermal plates These flattened invaginated sacs spread and finally fuse, thus separating the larval ectoderm and the thin amnion from the newly invaginated ectoderm of the larval worm within The larva completes development with the formation of the anus and other organs It then sheds the larval ectoderm and amnion and emerges as a young worm

Asexual reproduction and regeneration. Certain nemertineans have the ability to multiply asexually by fragmentation of the body Each fragment containing a part of the lateral nerve cords can regen

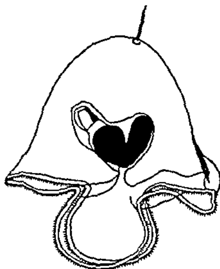


Fig 5 Pilidium of a nemertinean (From O Burger *Nemertini (Schnurwurmer), in H G Bronn ed Klassen und Ordnungen des Thier Reichs, vol 4,* 1903)

erate into a complete worm Other nemertineans, when irritated by handling or unfavorable conditions, fragment the body or evert the proboscis so that it breaks off The anterior region, including the foregut, can regenerate a new proboscis and a new posterior end In the process of anterior regeneration the wound is covered by migrating epidermal

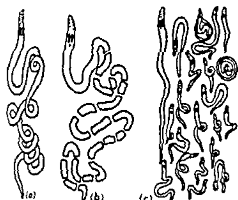


Fig 6 (a-c) Fragmentation in *Lineus vegetus* (From W R. Coe *Revision of the nemertean fauna of the Pacific coasts of North, Central, and northern South America* Allan Hancock Pacific Expeditions, vol 2, University of Southern California Press, 1940)

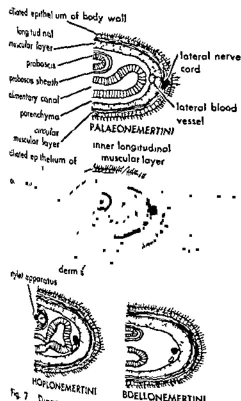


Fig. 7 Diagram of transverse sections of the body in the four orders of nemertineans, showing the arrangement of the muscular layers and the position of the lateral nerve cords and lateral blood vessels (From W R. Coe, *Biology of the nemerteans of the Atlantic coast of North America*, Trans Conn Acad Arts Sci. 35 145, 1943)

cells. Then, in the mass of mesenchyme cells which forms below the closure, three groups of cells appear, two lateral groups reforming the cerebral ganglia and a median group reforming the proboscis. Regeneration of the posterior end occurs by a lengthening of the body through the differentiation of mesenchyme cells *See* REGENERATION (BIOLOGY), REPRODUCTION, ANIMAL.

Ecology. The nemertineans are mostly marine, bottom dwelling worms, found in greatest numbers along the coasts of north temperate regions. They live under stones, among the tangled masses of plants, in sand, mud, or gravel, and sometimes form mucus-lined tubes A few are pelagic, fresh water, or terrestrial Certain species are commensal with other animals, but none may be regarded as parasitic in a strict sense

Phylogenetic relationships. The Rhynchocoela are related to the Platyhelminthes and probably have evolved from the same ancestral stock which gave rise to that phylum They resemble the flatworms in having the region between the epidermis and the gut filled with mesenchyme, in the arrangement of the nervous system, in the structure of the eyes, in the occurrence of ciliated grooves on the head and in showing spiral cleavage The pro- [illegible] the cerebral organs may be regarded as

[illegible]

That the Rhynchocoela represent the most highly organized acoelomate animals is indicated by the circulatory system, the presence of an anus, and the specialization of the epidermis All groups of animals more complex than the nemertineans have some kind of cavity a pseudocoele or coelom, between the body wall and the gut, instead of solid mesenchyme *See* COELOM, PSEUDOCOELOMATA

Classification. The phylum Rhynchocoela, containing about 550 known species is divided into 2 classes In the class Anopla the mouth is posterior to the brain the nerve cords lie under the epidermis or in the muscle layers of the body wall, and the proboscis is unarmed The two orders of the Anopla are separated on the basis of the musculature and the character of the dermis which is a connective tissue lying under the epidermis The order Palaeonemertini has the muscles arranged in two or three layers, if three, the innermost layer consists of circular fibers The dermis is gelatinous The order contains such genera as *Tubulanus*, *Carinoma*, and *Cephalothrix* In the second order, the Heteronemertini, the muscles are in three layers with the innermost longitudinal, and the dermis is fibrous The order contains, among others, the genera *Lineus*, *Cerebratulus*, and *Micrura*

The second class the Enopla has the mouth anterior to the brain the nerve cords internal to the muscles, and the proboscis often armed with stylets Of the two orders the first, the Hoplonemertini, has a proboscis armed with one or more stylets and has a straight intestine with paired lateral diverticula Representative genera are *Emplectonema* *Carci*

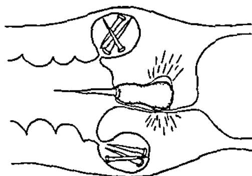


Fig 3 Stylets of *Amphiporus griseus* (From W R Coe Biology of the nemerteans of the Atlantic coast of North America Trans Conn Acad Arts Sci 35 282, 1943)

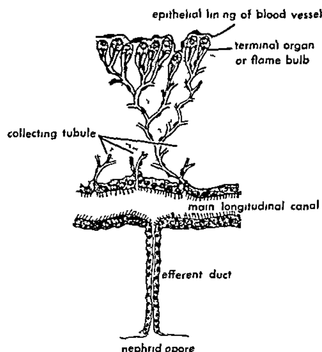


Fig 4 Diagram of the multiple protonephridium of a nemertinean (From W R Coe Unusual types of nephridia in nemerteans Biol Bull 58(3) 214 1930)

called athrocytes may surround the flame bulbs and tubules and are assumed to be excretory since they readily take up vital stains

The nervous system has a pair of cerebral ganglia forming the brain as well as two longitudinal nerve cords and many smaller nerves The ganglia and lateral cords may contain unusually large neurochord cells In the epidermis there are scattered sensory nerve cells probably tactile Chemotactile organs are situated in a pair of anteriorly placed cephalic grooves or in a flask shaped protrusible frontal organ A few to many simple eyes or ocelli may be present in front of the cerebral ganglia Statocysts the organs of equilibrium are rare Two blind canals the cerebral organs are invaginated from the epidermis and are closely associated with the cerebral ganglia These organs probably chemosensory open through pores in the cephalic grooves or on the body surface

There are no special respiratory organs and respiration occurs through the body surface

Nemertineans are usually either male or female but a few individuals have both sex organs. The ovaries or testes open by short ducts to the exterior Fertilization occurs outside the body in many species but may be internal in certain forms

Embryology. Cleavage of the fertilized egg or zygote, is spiral, with the cells of the lower quartet of the 8 cell stage rotated slightly, to lie in the furrows between the cells of the upper quartet Development is determinate, with the potentialities for future development of the embryo determined or fixed in the zygote before cleavage begins Isolated cells of the 2 cell stage result in dwarf larvae Isolated cells of later stages result in deficiencies *See* CELL LINEAGE, CLEAVAGE EMBRYONIC

After early cleavage the development in certain nemertineans may be direct that is, without a larva the embryo emerging from the egg membranes as a minute ciliated worm In others, the gastrula becomes a free swimming helmet shaped ciliated pilidium, formed by the downward growth of two ciliated lobes at the sides of the mouth and having an apical tuft of cilia In still other nemertineans the gastrula remains inside the egg membranes and becomes an oval ciliated larva known as Desor's larva which lacks the apical tuft and the oral lobes

Both the pilidium and the Desor's larva metamorphose into an adult worm by means of the invagination of seven or eight ectodermal plates These flattened invaginated sacs spread and finally fuse thus separating the larval ectoderm and the thin amnion from the newly invaginated ectoderm of the larval worm within The larva completes development with the formation of the anus and other organs It then sheds the larval ectoderm and amnion and emerges as a young worm

Asexual reproduction and regeneration Certain nemertineans have the ability to multiply asexually by fragmentation of the body Each fragment containing a part of the lateral nerve cords can regen

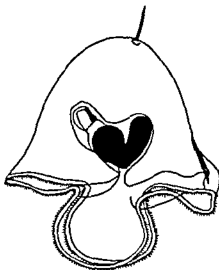


Fig 5 Pil dium of a nemertinean (From O Burger Nemertini (Schnurwurmer), in H G Bronn ed Klassen und Ordnungen des Thier-Reichs, vol 4, 1903)

er lo e nature of rhyolitic eruptions reduces greatly the amount of rhyolite formed as lava flows and correspondingly increases the amount formed as fragmental or tuffaceous material Great quantities of rhyolite may have been erupted as volcanic ash and dust and later incorporated in the thick sedimentary accumulations of geosynclines *See* Tuff

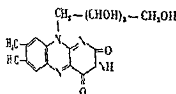
Rhyolitic magma may form by differentiation from basaltic magma perhaps in combination with more or less assimilation of siliceous material (sedimentary and granitic rocks) It may also be generated by melting portions of the earth s sialic la er *See* Magma Petrographic province.

[C A CA]

Riboflavin

Also known as vitamin B. Although widely distrib ' 'l in milk egg

b ing tructural formula



Riboflavin

It is stable to acid and oxidation but is rapidly de- stroyed by alkali at elevated temperatures and by

12 hours.

Riboflavin is determined by microbiological

glossitis corneal vascularization ac

usually paralleling metabolic activity The flavin functions biochemically in two coen

reactions by accepting hydrogen ions (protons) and electrons from one substrate and transferring them to another Over 14 riboflavin containing enzymes have been studied *See* Coenzyme, Enzyme

Most of the energy yielding oxidations of food to carbon dioxide and water occur through the cytochrome system in which the flavoprotein cytochrome reductase is essential L Amino oxidase, another flavoprotein is important in the interconversion of amino acids to nonnitrogenous metabolites of carbohydrate and fat metabolism *See* Cytochrome.

Riboflavin requirements do not appear to be related to caloric requirements or muscular activity but are affected by heredity growth environment age and health *Evidence in animals suggests a* need for increased riboflavin in low protein diets because of a decreased ability of the liver to retain the vitamin There is also evidence that as a result of intestinal synthesis by bacteria less riboflavin is required on diets high in carbohydrate than on those high in fat Human requirements for riboflavin are based primarily on urinary excretion data and studies in which riboflavin deficiency has been experimentally produced Riboflavin allowances of the National Research Council have been computed from protein allowances in such a way that 0 025 mg of riboflavin have been recommended per gram of dietary protein Thus daily riboflavin allowances vary from 0 4 mg for infants to 2 5 mg for 16- to 20-year old boys and lactating women *See* Vitamin [S N G]

Industrial production Riboflavin production is commercially accomplished by either direct chemical synthesis or fermentation

Chemical synthesis Industrial syntheses of riboflavin generally proceed along the lines of the original Karrer method with modifications by individual producers A typical manufacture starts with D-ribose which is condensed with 1 3 4-xylidine and simultaneously hydrogenated to form N D-ribitylxylidine The ribitylxylidine is coupled with diazotized aniline to produce 1 2 dimethyl 4-D ribitylamino 5 phenyl azo benzene This azo compound is hydrogenated to 1 2 dimethyl 4-D ribityl amino-5 aminobenzene and then is condensed with a mixture of alloxan and alloxantin to riboflavin By another method the azo compound is condensed directly with barbituric acid [L A F]

Fermentation Riboflavin production is also accomplished commercially by fermentation utilizing the synthetic ability of bacteria yeasts or fungi Production in the United States was 140 000 lb in 1957 with sales of $2 187 000 for medical use and for enrichment of human food 265 000 lb was produced for use in animal feed with sales of $3 426 000

for production of ri

dried on sand and after one or two carried through one or two tank inoculum stages

The final fermentation is carried out in tanks with a capacity of 10 000–100 000 gal with a me

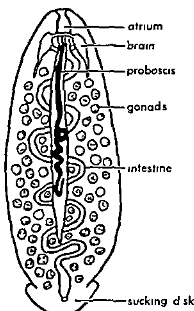


Fig 8 Diagram of *Malacobdella grossa* (From W R Coe *Biology of the nemerteans of the Atlantic coast of North America Trans Conn Acad Arts Sci*, 35 308, 1943)

nonemertes (on the gills and egg masses of crabs) *Amphiporus, Tetrastemma Prostoma* (in fresh water) *Geonemertes* (on land near the sea in tropical and subtropical regions) and *Nectonemertes* (pelagic) The second order the Bdellomorpha (Bdellonemertini), has an unarmed proboscis a sinuous intestine without diverticula and a posterior adhesive disk The only genus *Malacobdella* is commensal chiefly in the mantle cavity of marine clams where it feeds on plankton brought in by the ciliary current *See* ANOPLA, ENOPLA [A G H]

Bibliography L H Hyman *The Invertebrates* vol 2 1951

Rhyolite

A light colored aphanitic (not visibly crystalline) rock of volcanic origin composed largely of alkali feldspar and free silica (quartz, tridymite or cristobalite) with minor amounts of dark colored (mafic) minerals (biotite hornblende or pyroxene) If sodic plagioclase exceeds the amount of alkali feldspar the rock is a quartz latite Rhyolite *is chemically equivalent* to granite Quartz latite is chemically equivalent to granodiorite

Texture. The high silica content gives rhyolitic lava a high viscosity which hinders crystallization and promotes the formation of glass Rhyolites composed almost entirely of glass are called obsidian pitchstone perlite or pumice If the glass carries abundant scattered crystals the rock is a vitrophyre Typical rhyolites are mixtures of microcrystalline aggregates and glassy material but many are entirely crystalline and some grade into microgranite Porphyritic varieties are those in which numerous large crystals (phenocrysts) are disseminated throughout an aphanitic matrix The phenocrysts are usually quartz and sanidine Plagioclase if present, also occurs as phenocrysts and is usually zoned oligoclase In quartz latite, plagiocla (oligoclase or andesine) is the most common ph nocryst Quartz phenocrysts are well formed (e hedral), bipyramidal crystals but they common show corrosive effects including deep embayme and rounded crystal corners As the abundance phenocrysts increases the rock passes into porphyr *See* PORPHYRY, VOLCANIC GLASS

The finer features of rhyolite must be studied m croscopically The crystalline portion of the groun mass consists of quartz and tridymite more or le intergrown with alkali feldspar Cristobalite is le abundant Two types of rhyolite are recognize *normal* (potassic) and *alkali* (sodic) In the pota sic type the principal feldspars are sanidine ar orthoclase, in the sodic type the principal fel spars are soda rich (soda sanidine and albite) No mal rhyolite most commonly carries flakes of brow biotite mica which may be so strongly resorbed b the lava that nothing remains but patches of dus iron oxide which outline the earlier crystals Gree hornblende much as resorbed crystals is less abu dant, augite if present appears typically as phen crysts In alkali rhyolite the mafics are soda ric and generally confined to the matrix as patches riebeckite arfvedsonite hastingsite or aegirite D opside phenocrysts with rims of aegirite or aegirine augite may also occur Accessories almost unive sally present are zircon sphene magnetite an apatite Hematite as dusty material in the feldspa may be sufficiently abundant to color the rock rec

Structure Many rhyolites are streaked o banded owing to flowage of solidifying lava Th fluidal structures consist of curving or wavy trail of tiny crystals or of alternating layers of glass an microcrystalline material Spherulitic structure are characteristic of rhyolite These spheroida bodies consist of radial aggregates of alkali feld spar needles sometimes with intergrown quartz tridymite or cristobalite Spherulites may coalesce or may be arranged in trains parallel to fluida structure Tiny cavities within spherulites (litho physae) may be common The high volatile content of many rhyolitic lavas promotes the formation

physae *See* IGNEOUS ROCKS OBSIDIAN

tiny cracks or around phenocrysts and gradually spreads until the whole mass is converted to a minutely crystalline aggregate of quartz or tridy mite and alkali feldspar

Occurrence and origin Next to basalt and andesite rhyolite is considered the most common volcanic rock It commonly forms lava flows and tuffs Some small dikes and sills or the chilled mar gins of larger bodies may be composed of rhyolite The rock is widespread and occurs in large quanti ties associated with andesite basalt and small amounts of dacite and quartz latite The extremely

and other countries of the Middle and Far East rice is more important than wheat as a source of carbohydrates (see CARBOHYDRATE) In some countries of the Orient the consumption of rice per capita is estimated at 200–400 lb/year In contrast the yearly per capita consumption of rice in the United States is only about 5 lb The most important rice producing countries are India China Pakistan, and Indochina, but in many of the smaller countries rice is the leading food crop Although the acreage planted in rice is only about one half that planted in wheat, the total world production is nearly as great, due to higher average acre yields (see WHEAT)

Production and economic importance In the United States rice is produced on approximately 1,500,000 acres in contrast to over 60 000 000 acres of wheat, largely concentrated in selected areas of Arkansas, Texas Louisiana and California Although this represents less than 1% of the world rice acreage, the United States exports about 15% of its rice largely to Japan Korea and South American countries. The annual value of the rice crop in the United States for the period 1945–1954 was $209 019 900

Uses. Although eaten mainly as boiled rice a considerable amount of rice is consumed as breakfast cereals Rice starch also has many uses. Broken rice is used as a livestock feed and for the production of alcoholic beverages (see ETHYL ALCOHOL) The bran from polished rice is used for livestock feed the hulls are used for fuel and cellulose (see CELLULOSE) The straw is used for thatching roofs in the Orient and for making paper mats, hats and baskets Rice straw is also woven into rope and used as cordage for bags This crop serves a multitude of purposes in countries where agriculture is dependent largely upon the culture of rice

Persia and Egypt nearly 5000 years ago In the United States, rice was not grown until about 1685 In the latter part of the nineteenth century Louisiana became an important rice producing state early in the twentieth century rice production spread to south eastern Texas eastern Arkansas and north central California Rice is a comparatively new crop in its present area of greatest production

Unlike many other cereal grains, all cultivated varieties of rice belong to the same species and have 12 pairs of chromosomes as do most wild

the chromosome number is the same many of the ancient types have become so widely differentiated that hybrids between them are only partially fertile. See GENETICS

Rice is an annual grass plant varying in height from 2 to 6 ft (see ANNUAL PLANTS) Plants tiller that is develop new shoots freely the number depending upon spacing and soil fertility Among the many types grown some mature in 80 days others require over 200 days The inflorescence is an open panicle (Fig 1) Flowers are perfect and normally selfpollinated with some varieties exhibiting 3–4% natural crossing See FLOWER (BOTANY) INFLORESCENCE REPRODUCTION PLANT A distinct characteristic of the flower is six rather than the customary three anthers as in other grasses Spikelets have a single floret lemma and palea completely enclosing the caryopsis or fruit which may be yellow red brown or black See FRUIT (BOTANY) Lemmas may be awnless partly or fully awned (see GRASS CROPS) In threshing the caryopsis is not freed from the glumes and is called rough rice or paddy Grain color unmilled may be white brown amber red or black and the grain may be long and slender to short and thick

Varieties In India alone over 8000 varieties have been recorded and in the Philippines over 3500 varieties are known In contrast only a few varieties are grown in the United States Cultivated varieties are classified as upland and lowland The upland varieties can be grown without irrigation and are relatively unimportant in acreage The lowland types are grown submerged in water for the greater part of the season (see IRRIGATION OF CROPS) In contrast to most plants rice can thrive

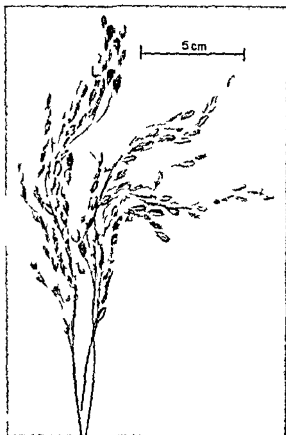


Fig 1 Panicles of rice Each spikelet has a single caryopsis enclosed in the lemma and palea

dium suitable for the organism being used For *Eremothecium* usually stillage (still slops) from the ethyl alcohol fermentation with skim milk, soy bean meal, or casein added is used as a proteinaceous source, a carbohydrate source such as maltose, sucrose, or glucose is added For *Ashbya*, commercially used media may contain corn steep liquor, and usually also some animal protein such as crude peptones, animal stick liquor, or fish stick liquor, the carbohydrate sources are similar to those used for *Eremothecium* The medium is aerated and usually also agitated during fermentation for 96–120 hours, optimum titers may be 3–6 g of riboflavin or more per liter *See* ASCOMYCETES, CARBOHYDRATE, CASEIN, ETHYL ALCOHOL, SOYBEAN

Riboflavin may be recovered for animal feed supplements by evaporation of the whole broth in multiple effect evaporators, followed by drum or spray drying For drug and fine food uses, pure crystalline riboflavin is isolated by heating the fermentation broth filtration, and precipitation of the riboflavin with dithionite (hydrosulfite) followed by several purification steps including crystallization

[R E B]

Bibliography H R Rosenberg, *Chemistry and Physiology of the Vitamins*, 1945, W H Schopfer *Plants and Vitamins*, 1943, L. A Underkofler and R J Hickey, *Industrial Fermentations*, vol 2, 1954

Ribonucleic acid

One of the two nucleic acids, also referred to as RNA A part of the RNA molecule is the sugar D ribose The other nucleic acid is deoxyribonucleic acid (DNA), whose molecule contains the sugar D 2 deoxyribose RNA, together with DNA and protein, is an essential component of all living matter *See* DEOXYRIBONUCLEIC ACID, NUCLEIC ACID

In the cell, RNA is usually found associated with proteins There are two main classes of RNA each of which is heterogeneous One is of low molecular weight (under 100 000), found in the cytoplasm, and termed soluble RNA, and one is of high molecular weight (over 1,000 000) and found associated with a roughly equal amount of protein in particles of 150–250 angstroms in diameter These are known as ribosomes, microsomal particles or ribonucleoprotein particles are found in all types of cells, and account for most of the cellular RNA They are either free in the cytoplasm or attached to various lipoprotein membranes They also may be present in the nucleus and possibly in mitochondria Highly polymerized RNA also occurs in many viruses, a number of which resemble ribosomes in size and composition *See* CYTOLOGY, VIRUS

RNA of all kinds is metabolically very active, it is continuously being broken down and resynthesized Its most important biological role seems to be concerned with protein synthesis in which it may function in several ways When existing RNA is damaged, or its synthesis is stopped or impaired, the cell immediately loses the ability to synthesize new protein Although the details are poorly un[derstood], the following picture of the roles of RN[A in] protein synthesis is a good working hypoth[esis]

Ribosomal RNA (or even whole ribosomes) [is] synthesized in the nucleus under "supervis[ion]" of the DNA in the chromosomes The DNA tr[ans]mits its genetic information to the polynucleo[tide] chains of RNA by controlling the base sequence [or] structure of the latter The ribosomes, now [pos]sessing the plan making proteins then move into the cytoplasm which provides the necess[ary] machinery for translating the plans into act[ual] proteins At this point, DNA itself becomes [dis]pendable, except for providing new ribosomes [as] needed *See* CELL (BIOLOGICAL), PROTEIN

When a cell is infected with an RNA virus [or] isolated infectious RNA from a virus, the vi[ral] RNA is thought to substitute for the riboso[mal] RNA and usurp the cellular machinery to m[ake] viral rather than host protein This change in [the] nature of the protein being made constitutes [the] most direct evidence that RNA can convey gene[tic] information

The "machinery" provided by the cytoplasm [in]cludes so called amino-acid activating enzy[mes] which change amino acids, which as such [are] chemically inert, so that they react easily w[ith] each other to form proteins These amino aci[ds], of which there are 20 or more different kinds, mu[st] however react in a strictly controlled sequence [in] order to yield specific proteins The ribosom[al] RNA often referred to as a template is thought [to] determine this sequence by virtue of its unique [ar]rangement of nucleotides, providing a three-dime[n]sional pattern such that only one given amino-ac[id] molecule can fit into any given space along t[he] template Until all adjacent spaces are filled a[nd] the peptide chain can be completed, the ami[no] acids must remain in an activated state This mea[ns] they must be transferred from the activating e[n]zymes to the template in a chemically bound fo[rm] (by means of some carrier) and must also be th[en] bound to the template The carrier should be ab[le] to recognize the proper position for the amino ac[id] it is carrying and get it close to the templat[e] After the transfer is completed it should be fr[ee] to pick up another charge This service is probab[ly] provided by the soluble RNA there being a specif[ic] RNA molecule for each kind of amino acid *Se[e]* AMINO ACIDS, ENZYME

The steps involved in the transfer from solubl[e] to ribosomal RNA have not been elucidated, bu[t] among the cofactors needed is GTP a triphosphat[e] derivative of the RNA constituent guanylic acid This might be provided by a reversible phospho[]rolysis of some kind of RNA thus pointing to [a] possible third function for RNA that of a reservoi[r] of cofactors for protein synthesis [I D R]

Rice

The plant *Oryza sativa* is the major source of hu[]man food for nearly one-half of the world's popula[]tion In China, Japan Korea the Philippines India

and other countries of the Middle and Far East rice is more important than wheat as a source of carbohydrates (see CARBOHYDRATE) In some countries of the Orient, the consumption of rice per capita is estimated at 200–400 lb/year In contrast the early per capita consumption of rice in the United States is only about 5 lb The most important rice producing countries are India China Pakistan, and Indochina, but in many of the smaller countries rice is the leading food crop Although the acreage planted in rice is only about one-half that planted in wheat, the total world production is nearly as great, due to higher average acre yields (see WHEAT)

Production and economic importance In the United States rice is produced on approximately 1,x00,000 acres, in contrast to over 60 000 000 acres

rice acreage, the United States exports about 1 5% of its rice largely to Japan Korea and South American countries The annual value of the rice crop in the United States for the period 1945–1954 was $209 019 900

Uses. Although eaten mainly as boiled rice a considerable amount of rice is consumed as breakfast cereals Rice starch also has many uses Broken rice is used as a livestock feed and for the production of alcoholic beverages (see ETHYL ALCOHOL) The bran from polished rice is used for livestock feed the hulls are used for fuel and cellulose (see CELLULOSE) The straw is used for thatching roofs in the Orient and for making paper mats, hats and baskets Rice straw is also woven into rope and used as cordage for bags This crop serves a multitude of purposes in countries where agriculture is dependent largely upon the culture of rice.

Origin and description Rice probably originated in the tropical climate of southern India and spread eastward into China and westward into Persia and Egypt nearly 5000 years ago In the United States rice was not grown until about 1685 In the latter part of the nineteenth century Louisiana became an important rice producing state Early in the twentieth century rice production spread to south eastern Texas eastern Arkansas and north central California Rice is a comparatively new crop in its present area of greatest production.

Like many other cereal grains all cultivated varieties of rice belong to the same species and have 12 pairs of chromosomes as do most wild types. The extent of variation in morphological and physiological characteristics within this single species is greater than for any other cereal crop Although the chromosome number is the same many of the ancient types have become so widely differentiated that hybrids between them are only partially fertile See GENETICS

Rice is an annual grass plant varying in height from 2 to 6 ft (see ANNUAL PLANTS) Plants tiller that is develop new shoots freely the number depending upon spacing and soil fertility Among the many types grown some mature in 80 days, others require over 200 days The inflorescence is an open panicle (Fig 1) Flowers are perfect and normally selfpollinated with some varieties exhibiting 3–4% natural crossing See FLOWER (BOTANY), INFLORESCENCE REPRODUCTION PLANT A distinct characteristic of the flower is six rather than the customary three anthers as in other grasses Spikelets have a single floret lemma and palea completely enclosing the caryopsis or fruit which may be yellow red brown or black See FRUIT (BOTANY) Lemmas may be awnless partly or fully awned (see GRASS CROPS) In threshing the caryopsis is not freed from the glumes and is called rough rice or paddy Grain color unmilled may be white brown amber red or black and the grain may be long and slender to short and thick

Varieties In India alone over 8000 varieties have been recorded and in the Philippines over 3500 varieties are known In contrast only a few varieties are grown in the United States Cultivated varieties are classified as upland and lowland The upland varieties can be grown without irrigation and are relatively unimportant in acreage The lowland types are grown submerged in water for the greater part of the season (see IRRIGATION OF CROPS) In contrast to most plants rice can thrive



Fig 1 Panicles of rice Each spikelet has a single caryopsis enclosed in the lemma and palea

when submerged because oxygen is transported from the leaves to the roots Rice varieties also are classified as common or glutinous according to starch characteristics Common rice is more extensively grown than glutinous rice which is used mainly in the Orient to prepare pastries and confections The grains of the widely grown common rice may be translucent to opaque hard to soft and when boiled may be soft to firm retaining their shape Thus among the thousands of varieties known there exists a wide range in plant and grain characteristics

Cultural practices Because of the peculiar conditions of growing a crop submerged in water rice land seldom becomes a part of a regular rotation system Often two or more successive crops of rice are grown and the land is then pastured or fallowed to control weeds Rice is grown on heavy soils underlain by an impervious subsoil to prevent seepage of water *See* SOIL

In the Orient and in many other countries rice fields are established by transplanting seedlings from seedling beds when the plants are 30–50 days old Fields to be transplanted are flooded and worked into a soft mud Clumps of three to four seedlings are pushed into the mud in rows to permit hand cultivation for control of weeds This sys



Fig 3 Hoja blanca d sease of rice (Inter Amer Inst tute of Agricultural Sc ences)

sized seedling bed is being grown Yields from transplanted rice usually exceed those planted directly from seed

In the United States the oriental system of transplanting is impractical because of labor costs Consequently rice in this country is seeded with a grain drill at 80 lb/acre prior to irrigation In California the seed often is sown from airplanes and left on the surface of the soil The field is then submerged prior to germination In the southern states fields are submerged with 1 2 in of water when the seedling plants are 6–8 in tall Later the depth of water is increased to 4–6 in and remains at that depth until the land is drained prior to harvesting the crop

The rice crop in the Orient and some other countries is harvested by hand (Fig 2) In the United



Fig 2 Hand harvest ng is the common pract ce used in many countr es

States rice is often harvested with a grain bin[illegible] in the southern states the bundles are shocked [illegible] the field and threshed when the grain moistu[illegible] content is reduced to 14% When harvested wit[illegible] combine a practice common in California t[illegible] grain should not be drier than 23–28% (*see* AG[illegible] CULTURAL MACHINERY) Rice dried to a lower mo[illegible] ture content while standing in the field may shatt[illegible] and the grain quality may be lowered When ha[illegible] vested at this high moisture content the grain [illegible] dried artificially to 14% moisture care being take[illegible] to keep the drying temperature below 110°F [illegible] avoid damage to the grain *See* AGRICULTURAL SC[illegible] ENCE (PLANT) GRAIN CROPS [I J J]

Diseases of rice Rice diseases are of great im portance because large numbers of people in Asi[illegible] depend mainly on this cereal for carbohydrate[illegible] Since population and rice production are in suc[illegible] close balance in countries like India China an[illegible] Japan relatively small losses can lead to cond tion[illegible] approaching famine In 1934 famine in certain di[illegible] tricts of Japan was caused by the rice blast disease In 1943 the famine in Bengal India was attribute[illegible] primarily to the loss in rice production caused b[illegible] Helminthosporium leaf spot disease

The major diseases of rice are the Helmintho sporium leaf spot disease blast or rotten neck caused by the fungus *Piricularia oryzae* (*se*[illegible] FUNGI) and a very recently discovered di ease th[illegible] hoja blanca of Cuba and Venezuela (*see* PLANT [illegible] [illegible]) The hoja blanca is believed to be ca ised by a vir is which is spread from di eased to health[illegible]

plants by an insect (Fig 3) Losses in susceptible varieties caused by this disease have been reported to be as high as 60% of the crop Control in these countries has not been satisfactorily obtained because presently available resistant varieties do not yield well enough for use

Another disease of rice is the bakanae disease of Japan caused by *Gibberella fujikurae* It is of particular interest, since this fungus is the producer of gibberellin, the plant growth stimulator, which has been recognized as a possible means of increasing plant growth and crop yields *See* GIBBERELLIN, PLANT DISEASE. [S J P C]

Processing. The four main parts of the rice kernel are the hull bran germ, and endosperm The purpose of milling rice is to separate the outer portions from the inner endosperm with a minimum of breakage. The various steps followed in rice milling are illustrated in the flow diagram (Fig 4)

Polished rice Rough rice, or paddy rice, as it is known, is separated from foreign material by vibrating sieves and air currents Various sizes of sieves separate seeds that are larger or smaller than rice and air currents carry off chaff, dust, and other

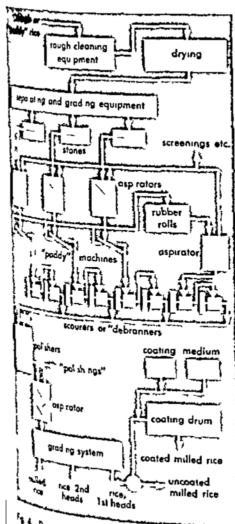


Fig 4. Diagram illustrating the processing of rice

lightweight material Rotating vertical cylinders containing indentations or perforations are also used to lift out and remove certain types of foreign

consist of two steel plates usually 4 ft or more in diameter, mounted horizontally, with the inner surface of the plates lined with a mixture of cement and coarse carborundum One plate rotates, and the other, set at the proper distance to permit rice grains to assume vertical positions, remains station

out breaking the endosperm After this first operation approximately 20% of the rough rice remains unhulled and must be processed further

The mixture of loose hull bran and germ as well as the unhulled kernels is conveyed to a stone reel This consists of a large revolving octagonal framework covered with wire screens where fine material known as stone-reel bran is recovered The large pieces of hull remaining are removed by suction in an aspirator and the mixtures of hull and unhulled grains left are separated in a paddy machine, a large box shaker fitted with vertical plates to form zigzag divisions which separate the lighter unhulled paddy grains from the heavier hull grains The plates and the shaking action cause the paddy grains to move gradually upward and over the machine into a trough while the heavier hulled grains are collected at the lower side The unhulled grains are sent through auxiliary hulling stones, which are set to smaller clearances, from there, they reenter the main stream going to the stone reel

Rice with hulls removed is called brown rice because it retains a light brown color from the bran coat Removal of the bran layers is done in two sets of hullers generally referred to as first or second break operations The term huller is a misnomer since the function of this machine is to remove bran layers instead of hulls The bran layers are removed by a scouring action between the inner walls of the huller and the rapidly revolving grooved inner core The mixture of scoured kernels and bran passes through the first break bran reel where the bran is separated Then the rice goes to the second break hullers Similar to the first but adjusted for closer action these remove the remaining part of the bran coat After leaving the second set of hullers, the rice has lost much of its brown color And it loses still more in the second bran reel

Pearling cones are used in some mills as an adjunct to the hullers or in place of the last set of hullers A pearling machine consists of a cone, coated with abrasive material, which revolves inside a heavy wire screen The scouring action can be adjusted to suit requirements

Rice passes from the second break huller or from the pearling cones, to a cooling bin and then

to a brush machine The latter consists of a vertical, cylindrical frame covered with soft leather strips revolving at a high rate of speed within a cylinder of wire screening The rice is rubbed smooth by friction during its travel through the machine A considerable amount of heat is generated in a brush machine, and cooling is accomplished by a stream of air which travels countercurrent to the flow of the rice From the brush machine, the rice passes through the brewers' reel, where the smallest fragments are removed This reel usually is covered with a mesh screen of proper size to remove the rice fragments, known as brewers' rice because this material has a ready market in the manufacture of beer

Often a coating of glucose or talc is applied in revolving tumblers These are cylinders about 9 ft long and 4 ft in diameter, set on an incline of about 15° from the horizontal Not all rice is coated with glucose or talc In some markets an uncoated grade of rice is preferred, but the luster of the grains is greatly improved by coating

Clean rice is graded by machines which consist of rows of vertical disks mounted on revolving shafts The indented disks collect the smaller sizes carry them over, and discharge them into a chute leading to cylinders where further separations take place The kernel size of the separations desired can be changed by raising or lowering the edge of the apron on the stationary axis of the machine

Many classes and grades of rice are marketed The by products of rice processing, such as hulls and bran, are generally sold for animal feed Some rice hulls are used in the manufacture of purified alpha cellulose

The common yields of the various products produced from rough rice are as follows hulls, 20%, brown rice 79%, whole grains, 58%, three quarters to half grains, 3%, screenings 6% By products obtained are bran, 8%, and rice polishings, 2%

Converted rice In the course of manufacturing polished rice much of the nutritive value in both protein content and vitamins, is removed Specifically 76% of thiamine 57% of riboflavin, and 63% of niacin are lost for human consumption Because of this efforts have been made to resupply the vitamins by a patented converted rice process, and by the Malek parboiling process *See* NIACIN RIBOFLAVIN, THIAMINE

In the process of milling brown rice to white rice, there is a marked reduction in B vitamins and iron Since customers prefer well milled white rice to brown rice, there has been much interest developed in methods to retain more of the B vitamins in the milled rice This problem has been approached in two ways by removing less of the brown bran layers and germ in milling, and by processing the rice prior to milling in such a way as to transfer vitamins and other water soluble nutrients from the outer portions of the grain into the endosperm

The first method produces undermilled or unpolished rice The outer bran is removed but the inner layers retained Since the greatest decrease in B vitamins and minerals occurs in the first break only limited improvement in nutritive value can be effected by decreasing the degree of refinement in subsequent stages of the milling process Attention has therefore been focused on methods to increase the vitamin content of the milled rice by subjecting rough rice to a parboiling treatment prior to milling

Converted rice is produced by the following process rough rice is cleaned and placed in steeping tanks which are sealed and evacuated to remove air from the tank and the interspaces of the rice kernels Hot water is added and pressure is applied for a steeping treatment This transfers the water soluble vitamins from the bran and germ into the endosperm Then the rice is subjected to live steam dried under vacuum, and finally milled and polished Converted rice is a highly milled polished product of greatly improved nutritive value

[J A SH

Bibliography J B Reed, *The By Products of Rice Milling*, US Dept Agr Bull 570, 1917, T B Wayne, Modern rice milling, *Food Ind*, 2 492-495, 1930

Ricinulei

An order of extremely rare arachnids, also known as the Podogona, with a body less than 1 in length Superficially, they resemble ticks in general appearance and movement, and are found only in tropical Africa and in the Americas, from the Amazon to Texas The two anterior pairs of appendages are chelate The terminal segments of the third legs of the male are modified as copulatory structures Less than 20 modern species are known The occurrence of several fossils from Carboniferous time suggests that the group was formerly more common Virtually nothing is known about the reproduction, growth, and ecology of the Ricinulei *See* ARACHNIDA [C C H

Rickets

A disorder of calcium and phosphorus metabolism primarily affecting bony structures due to vitamin D deficiency Precursor substances from the diet are normally converted to vitamin D by the action of ultraviolet light (sunlight) on the skin Therefore infants are affected, especially in winter Dark skinned peoples require additional vitamin D as dietary supplement because their skin pigmentations interfere with natural production *See* CALCIUM METABOLISM, PHOSPHATE METABOLISM, VITAMIN D

In children rickets consists of defective calcification and excessive production of cartilage at the ends of growing bones Restlessness constant movement often with hair loss from pillow contact, and defects of the ribs and long bones are typical symptoms Bowlegs and pigeon breast may result, in addition to craniotabes, or softening of the flat skull bones These later develop abnormally to form square boxlike skull There is an increased suscep

ability to fracture delayed tooth eruption and enamel defects, and other indications of faulty mineral deposition. See TOOTH DISORDERS

In adults, vitamin D deficiency produces osteomalacia, or demineralization of bones It is seen in a softening of the spine pelvis, and leg bones Fractures and deformities due to compression of the defective bones are common In addition, low calcium levels in the blood may be present with resulting irritability spasms and convulsions particularly of the hands, face and larynx See BONE, BONE (BIOPHYSICS)

Vitamin D regulates the absorption of calcium and phosphorus from the gastrointestinal tract, thereby maintaining blood levels [E G S T]

Rickettsiales

An order of bacterialike organisms of the class Microtatobiotes Individual organisms are not ultramicroscopic except, perhaps in rare filterable

... cause of epidemic or louse borne typhus fever of man Another important species is the agent of the world wide Q fever of man and animals, Coxiella burnetii Although its reservoir is in ticks, it is most often spread by inhalation of contaminated dusts from various sources A canine disease of the U.S Pacific Coast caused by *Neorickettsia helminthoeca* is transmitted by parasitic intestinal trematodes See Q FEVER, TYPHUS FEVER EPIDEMIC (LOUSE BORNE)

Chlamydiaceae These agents of diseases related to trachoma and parrot fever are not insect borne and comprise 5 genera (*Chlamydia Colesiota Ricolesia, Colettsia* and *Miyagawanella*) and at least 20 well-described species The type species is *Chlamydia trachomatis* cause of the widespread eye disease, trachoma, while some of the species assigned to *Miyagawanella* cause psittacosis lymphogranulomatous venereal disease and virus pneumonia. Many agents of the family produce forms of animal conjunctivitis especially in Africa See LYMPHOGRANULOMA VENEREUM, PSITTACOSIS

Bartonellaceae Members of this family are parasites of the red blood cells of man and animals The family comprises 4 genera (*Bartonella, Grahamella, Haemobartonella* and *Eperythrozoon*) and at least 20 accepted species The type species is *Bartonella bacilliformis* the cause of both anemic disease (Oroya fever) and eruptive disease (Verruga Peruana) in man on the northwest coast of South America ... by sandflies of the



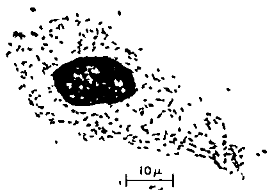
Rickettsia prowazekii, cause of epidemic typhus, in chick embryo, yolk sac cell, showing pleomorphism (Photography by N J Kramis Rocky Mountain Laboratory USPHS)

genus *Phlebotomus* Other species are found in native mice and other small rodents and occasionally in domestic animals See CARRION'S DISEASE

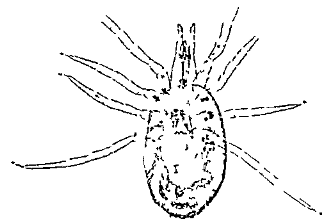
Anaplasmataceae These are parasites of the red blood cells chiefly of domestic cloven hoofed animals One genus *Anaplasma*, and three species are recognized The bacterial or protozoan nature of the *Anaplasma* species has been disputed but all are transmitted by ticks and mechanically by biting flies as well as by surgical instruments The type species is *Anaplasma marginale* cause of world wide malignant anaplasmosis of cattle See MICROTATOBIOTES [C B P]

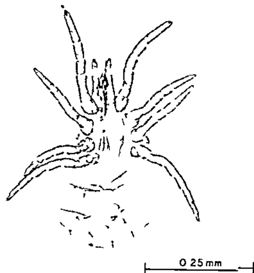
Rickettsialpox

A benign infectious disease similar to typhus, caused by bacterialike microorganisms *Rickettsia akari* It is transmitted by the mite *Allodermanyssus sanguineus* (see MESOSTIGMATA RICKETTSIALES) The disease was recognized in the late 1950s and has been reported only in Atlantic Coast cities of the United States in Africa and in European



Rickettsia akari causative agent of rickettsialpox in stained smear from yolk sac of infected chicken embryo (Photomicrograph by N J Kramis Rocky Mountain Laboratory, USPHS)





Allodermanyssus sanguineus, the mite vector of rickettsialpox (a) Adult female (b) Nymph (*From C B Philip in T T Mackie et al, Manual of Tropical Medicine 2d ed, Saunders, 1954*)

U S S R In all these areas strains of *R akari* have been recovered from house mice and their mites, as well as from patients, often in the environs of modern apartments Unlike trombiculids which transmit scrub typhus these mites are parasitic in older stages and leave the host between feedings In addition to a rash patients exhibit a primary eschar, or ulcer and often show a swelling of lymph glands similar to that in scrub typhus and fievre boutonneuse An agglutination test, the Weil Felix OX_{19}, is only occasionally positive Complement fixation cross immunity in guinea pigs and intranuclear growth of organisms show a relationship to spotted fever group rickettsioses, but so far no toxic factor has been demonstrated *See* RICKETTSIOSES

[C B P]

Rickettsioses

Infectious diseases affecting man and some animals caused by bacterialike microorganisms The human pathogens are members of the genus *Rickettsia* and the related genus *Coxiella* which are classified in the family Rickettsiales (*see* RICKETTSIALES) These rickettsial agents comprise a restricted group found in parasitic arthropods such as insects ticks and mites, which act as vectors are noncultivable on ordinary media are usually intracellular in growth, and usually will not pass medium to fine filters Dr H T Ricketts, who died of typhus fever in Mexico after confirming the causative agent, was memorialized in the naming of the group Related rickettsialike organisms which include harmless or beneficial symbiotes of invertebrate hosts or pathogens of vertebrate hosts only, most of them originally assigned to this genus, have recently been transferred to other genera

Morphology. The microorganisms are gram negative, small coccoid to rod shaped or filamentous, and often pleomorphic In size they generally lie between the bacteria and the filterable viruses and are visible under the light microscope only *Coxiella* has a readily filterable phase Under the electron microscope, magnified 10 000–100 000 times, they exhibit an outer envelope and an inner matrix which can be separated by suitable physicochemical techniques, the matrix appears to include numbers of dense granules

Growth and reproduction. They grow intracellularly by simple division, a report was made in 1957 of the cultivation of one agent in a cell free but intracellularlike environment However, this remains unconfirmed One other species grows extracellularly in the gut of its body louse host but will not grow in yolk sacs of embryonated chicken eggs as will the other *Rickettsia* Eventual cultivation on suitable artificial media is to be expected under present intensive research Though there is no spore formation, a few species are resistant to environmental stress and will live for a few months or even years in such detritus as dried arthropod feces or pulverized infected animal offal and fomites Others persist in tissues of the recovered animal



Coxiella burnetii the rickettsialike, causative agent of Q fever showing outer envelope and inner matrix in cells under the electron microscope (*Electron micrograph by R O Ormsbee and E Ribi, Rocky Mountain Laboratory USPHS*)

No complicated cycle of growth has been postulated. Organisms may be scattered diffusely through the cytoplasm of the cell or more compact colonies may occur such as the so called Mooser bodies ver group l nucleus

are the such human diseases as epidemic and endemic, or murine typhus the Rocky Mountain spotted fever group of diseases including fievre boutonneuse of the Old World and several tick typhuses such as tsutsugamushi disease or scrub typhus of the Far East as well as the rickettsialpox of the Atlantic Coast of the United States and of USSR Clinically all are characterized by fever and a cutaneous eruption or rash which is absent only in related Q fever In addition a primary ulcer or eschar at site of vector attachment and swollen glands which drain this area are associated with fievre boutonneuse subgroup Siberian and North Queensland tick typhuses rickettsialpox and scrub typhus Orchitis and swelling of the scrotum customarily occur in male guinea pigs infected with murine typhus and the spotted fever group *See* FIEVRE BOUTONNEUSE RICKETTSIALPOX TYPHUS SCRUB TYPHUS FEVER ENDEMIC (FLEA BORNE) TYPHUS FEVER EPIDEMIC (LOUSE BORNE)

Rickettsial study techniques The relationships and characteristics of these agents are still under investigation The techniques for studies of the rickettsiae include appearance and staining reactions under microscopic examination survival under different biological and physicochemical stresses such as in animal or insect wastes drying under vacuum and storage by freezing growth characteristics in chicken embryos tissue cultures and in unnatural arthropod hosts such as injected mealworms or in natural hosts such as intrarectally injected body lice or ticks fed artificially through membranes and glass capillaries differential behavior of suspensions in cellulose ion exchange columns and other chrom

with nonspecific Weil Felix tests and specific rickettsial toxin agglutination and complement fixation tests differential susceptibility of pathogenicity for and survival in various laboratory animals



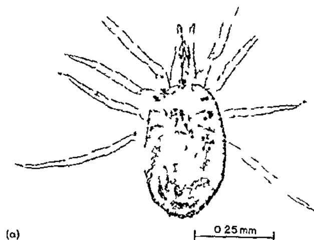
In a cellular rickettsiae under the light microscope (a) *Rickettsia typhi* of endemic typhus in characteristic infected Mooser cell from tunica of male guinea pig (from T T Mackie G W Hunter and C B Worth *A Manual of Tropical Medicine* 2d ed Saunders 1955) (b) *R tsutsugamushi* of scrub typhus in yolk cell of chicken embryo (from C B Philip Scrub typhus *Sci Monthly* 69 281–289 1949)

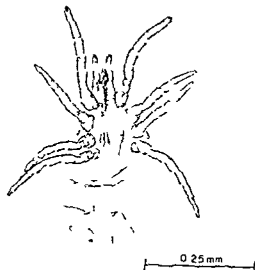
Staining characteristics These are usually determined by the Giemsa or Macchiavello methods Rickettsiae on air dried (lightly flamed) smears from infected yolk sacs or from appropriate animal tissue such as scrotal sacs of male guinea pigs or scrapings from body cavities of mice are colored purplish with Giemsa and bright red against a blue background with Macchiavello stains

Serology The Weil Felix test is based on production of nonspecific agglutinins (antibodies) in blood of patients by the agents of the typhus and spotted fever groups and of scrub typhus in blood of patients These react against the nonmotile O form of the bacterium *Proteus vulgaris* strain X The proteus strains OX_{19} and to a lesser extent OX_2 react against the typhus and spotted fever groups but do not differentiate between them while OXK is diagnostic for scrub typhus (*see* PROTEUS) A rise in titer in serially drawn specimens from patients after the fifth day of the disease has special significance Such agglutinins fade in a few weeks are entirely absent in cases of trench and Q fever and uncommon in rickettsialpox The complement fixing antibodies last longer and are also used to differentiate groups and in some cases species of *Rickettsia* as are specific agglutinins which cause clumping of organisms suspended in

the spotted fever group and two strains of tsutsugamushi disease *See* SEROLOGY

Immunology There appear to be five distinguishable groups of human pathogens namely the





Allodermanyssus sanguineus the m te vector of rickett sialpox (a) Adult female (b) Nymph (*From C B Philip in T T Mackie et al Manual of Trop cal Medicine 2d ed Saunders 1954*)

U S S R In all these areas strains of *R akari* have been recovered from house mice and their mites as well as from patients often in the environs of mod ern apartments Unlike trombiculids which trans mit scrub typhus these mites are parasitic in older stages and leave the host between feedings In ad dition to a rash patients exhibit a primary eschar or ulcer and often show a swelling of lymph glands similar to that in scrub typhus and fievre bouton neuse An agglutination test the *Weil Felix* OX_{19} is only occasionally positive Complement fixation cross immunity in guinea pigs and intranuclear growth of organisms show a relationship to spotted fever group rickettsioses but so far no toxic factor has been demonstrated *See* RICKETTSIOSES

[C B P]

Rickettsioses

Infectious diseases affecting man and some animals caused by bacterialike microorganisms The human pathogens are members of the genus *Rickettsia* and the related genus *Coxiella* which are classified in the family Rickettsiales (*see* RICKETTSIALES) These rickettsial agents comprise a restricted group found in parasitic arthropods such as insects ticks and mites which act as vectors are noncultivable on ordinary media are usually intracellular in growth and usually will not pass medium to fine filters Dr H T Ricketts who died of typhus fever in Mexico after confirming the causative agent was memorialized in the naming of the group Related rickettsialike organisms which include harmless or beneficial symbiotes of invertebrate hosts or patho gens of vertebrate hosts only most of them origi nally assigned to this genus have recently been transferred to other genera

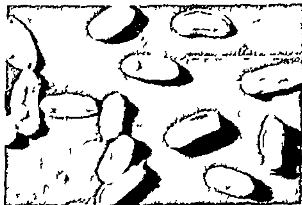
Morphology The microorganisms are gram negative small coccoid to rod shaped or filamen tous and often pleomorphic In size they generally lie between the bacteria and the filterable viruses and are visible under the light microscope, only *Coxiella* has a readily filterable phase Under the electron microscope magnified 10 000–100 000 times they exhibit an outer envelope and an inner matrix which can be separated by suitable physico-chemical techniques the matrix appears to include numbers of dense granules

Growth and reproduction They grow intra cellularly by simple division a report was made in 1957 of the cultivation of one agent in a cell free but intracellularlike environment However this re-mains unconfirmed One other species grows extra cellularly

not grow

as will the

suitable artificial media is to be expected under present intensive research Though there is no spore formation a few species are resistant to en vironmental stress and will live for a few months or even years in such detritus as dried arthropod feces or pulverized infected animal offal and fomites Others persist in tissues of the recovered animals



Coxiella burnetii the rickettsial ke causative agent of Q fever show ng outer envelope and inner matr x in cells under the electron m croscope (Electron m cro-graph by R O Ormsbee and E Ribi Rocky Mounta n Laboratory USPHS)

Each such elemental volume is located from a fixed point O by means of a position vector $\mathbf{r}$ and has a mass density ρ.

The total mass m of the rigid body can then be expressed by the following volume integral

$$m = \int_V \rho \, dV \tag{1}$$

The center of mass of the rigid body is then defined as the point within the rigid body whose position vector $\mathbf{r}_c$ is given by

$$\mathbf{r}_c = \frac{1}{m} \int_V \mathbf{r} \rho \, dV \tag{2}$$

The coordinates of the mass center are then

$$x_c = \frac{1}{m} \int_V x\rho \, dV \qquad y_c = \frac{1}{m} \int_V y\rho \, dV$$
$$z_c = \frac{1}{m} \int_V z\rho \, dV \tag{3}$$

where x, y, z are the coordinates of the volume dV

Translational motion. If a group of forces push or act on a rigid body in such a way that the line of action of these forces passes through the center of mass, pure translational motion results (Fig. 3) If the line of action does not pass through the center of mass the resulting motion is

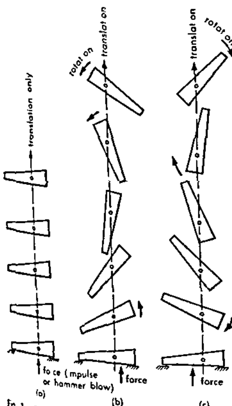


Fig. 1 Resultant motion for a force acting (a) at the center of mass, (b) right of the center of mass, and (c) left of the center of mass. The resultant motion in each case is shown

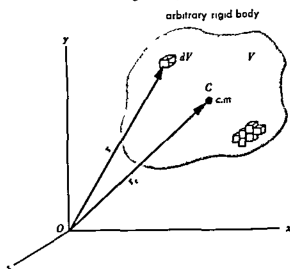


Fig 2 A rigid body of arbitrary shape may be considered to be composed of a continuous distribution of volume elements A typical volume element dV and position vectors are shown

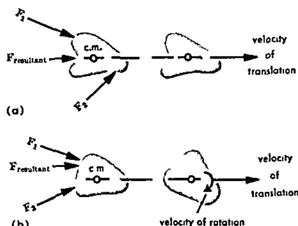


Fig 3 Illustration showing (a) pure translation caused by a resultant force passing through the center of mass of a rigid body and (b) combined translation and rotation caused by a resultant force which does not pass through the center of mass

a combination of both translation and rotation However in either case the motion of the center of mass of the rigid body is governed by the following equation, derived directly from Newton's second law applied to each particle of the rigid body and summed for all such particles

$$\Sigma \mathbf{F} = m\mathbf{a}_c = m \frac{d^2\mathbf{r}_c}{dt^2} \tag{4}$$

where $\mathbf{a}_c = d^2\mathbf{r}_c/dt^2$ is the vector acceleration of the center of mass and $\Sigma\mathbf{F}$ is the sum of all external forces acting on the rigid body All internal forces sum to zero because such forces between particles of the body occur in equal and opposite pairs (*Newton's third law*)

Equation (4) states that the motion of the center of mass of a rigid body is the same as would be

typhus group, the tsutsugamushi group the spotted fever group, Q fever, and trench fever For the last, there are no susceptible nonprimate laboratory animals Partial to complete cross immunity exists between these pathogens and is detailed under articles on the individual diseases

Vectors, treatment, and prevention. The rickettsial vectors the treatment for, and prevention of, rickettsioses are presented in this section

Vectors Most rickettsiae are transmitted by specific arthropods, such as body lice, fleas ticks, or mites, but a new species from U S S R *R pavlovskii* is reported in all of the last three See ACARINA, ANOPLURA, SIPHONAPTERA

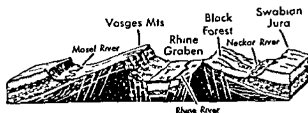
Treatment All these agents are susceptible to the so called broad spectrum antibiotics, including chloramphenicol and the tetracyclines, chlortetracycline and oxytetracycline, under appropriate dosages *See* ANTIBIOTIC, CHLORAMPHENICOL, CHLORTETRACYCLINE, OXYTETRACYCLINE, TETRACYCLINE

Prevention Protection of individuals is accomplished in two ways (1) control of the source of infection such as lice and flea vectors in the typhus fevers rat and mouse reservoirs in endemic typhus and rickettsialpox and contaminated environments, such as dust and milk, in Q fever, and (2) vaccination—for example practical vaccines of killed organisms have been commercially developed against epidemic typhus and Rocky Mountain spotted fevers In addition efficacy of live attenuated vaccine against the epidemic typhus has also been demonstrated [C B P]

Bibliography A J Rhodes and C E Van Rooyen *Textbook of Virology*, 3d ed 1958, T M Rivers and F L Horsfall Jr (eds) *Viral and Rickettsial Infections of Man*, 3d ed 1959

Rift valley

An elongated relatively narrow depression caused by the subsidence of a crustal block between two or more faults The boundary faults are steeply inclined down toward the downthrown block and where the direction of displacement has been ascertained it indicates that the dislocations are gravity faults Thus rift valleys are the surface expression of large graben The term rift is used by some geologists as a synonym for graben Others define a rift as a strikeslip fault that parallels the trend of the regional structure *See* FAULT AND FAULT STRUCTURES, GRABEN



A generalized cross section of the rift valley of the Rhine and the adjoining block mountains of the Vosges and the Black Forest (*Drawing by E Raisz from P E James An Outline of Geography, 2d ed Ginn, 1943*)

Rift valleys commonly have lengths measured in hundreds of miles with relief at their margins of hundreds or thousands of feet Rift valleys cut across broadly arched regions and are produced by the lateral extension of the rocks of these areas The association of basaltic lavas with many rift valleys suggests that the boundary faults are major breaks in the crust and pass downward into the subcrustal region of the earth Rift valleys, therefore, have been interpreted to be a part of the major tectonic pattern of the earth and to be the result of deep seated deforming pressures *See* AFRICA, TECTONIC PATTERNS [P H O]

Rigel

Beta Orionis, a blue supergiant of spectral type B8 Although Rigel is one of the apparently brightest stars in the sky (0 2 visual magnitude), it is too distant to have a measurable parallax or proper motion Its luminosity can be estimated as -7^m or -8^m because it has a faint companion with which it is probably physically connected So high a luminosity about 60 000 times that of the Sun, means that Rigel is an exceptionally young star in rapid evolution with a life span of only a few million years *See* STAR [J L G R]

Rigid body

An idealized extended solid whose size and shape are definitely fixed and remain unaltered when forces are applied Treatment of the motion of a rigid body in terms of Newton's laws of motion leads to an understanding of certain important aspects of the translational and rotational motion of real bodies without the necessity of considering the complications involved when changes in size and shape occur Many of the principles used to treat the motion of rigid bodies apply in good approximation to the motion of real elastic solids *See* RIGID BODY DYNAMICS [D W I]

Rigid-body dynamics

A rigid body is defined as an assemblage or system of mass particles that are located rigidly with respect to one another and therefore can have no motion relative to each other Motion of a rigid body can occur by movement of all points of the body in a parallel direction through equal distances during a given interval of time called translation or by movement of all points in circles about a common axis with a common angular velocity, called rotation or by combined translation and rotation

Center of mass The center of mass of a rigid body is a single point located within the body such that any force acting externally on the body along a line of action which passes through this point will result in pure translation, that is, no rotation (see Fig. 1) *See* CENTER OF MASS

In order to calculate the position of the center of mass consider a rigid body divided into elemental volumes (or particles) labeled dV in Fig 2

angular velocity ω of the body and the vector distance from the axis of rotation to the point in question.

$$v_B = \omega \times r = \frac{dr}{dt} \quad (8)$$

Or in this case,

$$v_B = \omega r \theta_1 \quad (9)$$

where v_B is the vector velocity of point B in a direction *normal to the radius* r *and angular velocity* vector ω (see Fig. 7). This direction is designated by the unit vector θ_1. The acceleration of point B is found by taking the derivative of v_B in Eq. (8), or

$$a_B = \frac{dv_B}{dt} = \frac{d}{dt}(\omega \times r) = \left(\frac{d\omega}{dt} \times r\right) + \left(\omega \times \frac{dr}{dt}\right)$$
$$= (\alpha \times r) + \omega \times (\omega \times r) \quad (10a)$$

or in the planar case of Fig. 7

$$a_B = \alpha r \theta_1 - \omega^2 r r_1 \quad (10b)$$

The acceleration of B is therefore composed of two parts—a tangential acceleration αr and a radial acceleration directed inward toward the axis of rotation equal to $\omega^2 r$.

For a three-dimensional rigid body, if the vector r is specified as the radius vector from any point on the axis of rotation to the point B, Eqs. (8) and (10a) are equally valid for this case. See ACCELERATION; VELOCITY; *see also* ROTATIONAL MOTION.

Rotation about translating axis. If a body rotates about an axis that is translating, for example, an automobile wheel or a spinning missile, the velocity of each point is the vector sum of the velocity of translation of the axis of rotation and the velocity

[illegible]

of a rigid body are again derivable from Newton's second law. In Fig. 2 a rigid body of arbitrary shape is located and moving in an inertial system. This body may be subdivided into mass particles dV of mass m_i and Newton's law applied to each particle.

$$m_i \frac{d^2r}{dt^2} = F \quad (11)$$

If both sides of Eq. (11) are multiplied by the position vector r (by means of a cross product)

$$m_i\left(r \times \frac{d^2r}{dt^2}\right) = (r \times F) \quad (12)$$

Here the vector product $(r \times F)$ is defined as the moment M or torque about the point O of the resultant force F acting on dV (see TORQUE). Equation (12) may be written as follows

$$m_i\left(r \times \frac{d^2r}{dt^2}\right) = m_i \frac{d}{dt}\left(r \times \frac{dr}{dt}\right) = \frac{d}{dt}\left(r \times m_i \frac{dr}{dt}\right)$$
$$= \frac{d}{dt}(r \times m_i v) \quad (13)$$

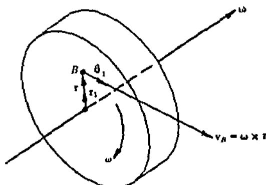


Fig. 7 Illustrating velocity of point B for a rotating disk

where r, v and m_i are the position vector, velocity and mass, respectively, of the element dV. Equation (13) is valid for each elemental particle of the rigid body and by summing the equation for all such particles of the body the equation of angular motion is obtained from the following

$$\Sigma M_o = (M_o)_{ext} = \frac{d}{dt}\int_V (r \times v)\rho\, dV = \frac{dH_o}{dt} \quad (14)$$

where m_i has been replaced by $\rho\, dV$ as before. In Eq. (14) the expression

$$\int_V (r \times m_i v) = \int_V (r \times v)\rho\, dV = H_o \quad (14a)$$

is called the moment of linear momentum or the angular momentum about a point O of a rigid body. On the left-hand side of Eq. (14), the summation process results *in zero for all equal and opposite* collinear forces between particles (internal forces), and only the moment M_o of the external forces exists.

Equation (14) expresses the principle of angular momentum, that the time rate of change of angular momentum of a rigid body about any point O fixed in an inertial system is equal to the resultant moment of external forces about the same point O. If the rigid body is not acted upon by any external moment, the angular momentum must remain constant. This is the principle of conservation of angular momentum.

An important extension of the angular momentum Eq. (14) is as follows. *Equation (14) is precisely valid for the case where the angular momenta and the moments of external forces are taken about the center of mass of a rigid body, even though the center of mass does not remain at rest in an inertial system.* In this case the equation is

$$M_c = \frac{dH_c}{dt} \quad (15)$$

where c denotes the center of mass. Equation (15) is useful, for example, in predicting the motion of a projectile moving through air under accelerated conditions. The angular position of the projectile affects the aerodynamic forces acting on the projectile, and Eq. (15) is necessary for the solution of the problem. The fact that the angular equa-

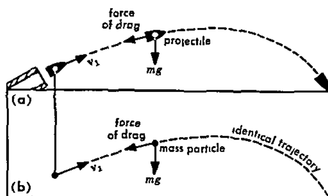


Fig 4 Translational motion of projectile The motion of the center of mass of the projectile in (a) is the same as the motion in (b) of a mass particle of equal mass and subjected to the same external forces

the case if all of the mass of the body were con centrated at this mass center and all external forces were applied there The motion of translation of the center of mass of a rigid body is therefore de termined by the same methods as those used to de termine the motion of a particle (see Fig 4) The one significant difference from particle dynamics, however, is that the forces are those acting on the actual rigid body and as such often depend upon the shape of the body, its orientation in space, and sometimes upon the time rate of change of this ori entation in space Examples are the motion of an airplane, dirigible, or projectile through air or of a submarine through water

The motion in translation may also be expressed in terms of a vector quantity called linear momen tum Because the velocity of the center of mass is given by

$$\mathbf{v}_c = \frac{d\mathbf{r}_c}{dt}$$

Eq 4 may be written

$$\Sigma\mathbf{F} = m\frac{d\mathbf{v}_c}{dt} = \frac{d}{dt}(m\mathbf{v}_c) = \frac{d\mathbf{P}}{dt} \qquad (5)$$

where $\mathbf{P} = m\mathbf{v}_c$ is defined as the linear momentum of the rigid body Thus the resultant external force $\mathbf{F}$ acting on a rigid body is equal to the time rate of change of the linear momentum of the body

If no external forces are acting Eq (5) states that the linear momentum of the body remains constant—a statement of the law of conservation of linear momentum See CONSERVATION OF MOMENTUM, MOMENTUM

Rotational motion. If a rigid body is rotating about an axis fixed in space, for example, a wheel turning on a shaft, all points on the axis remain fixed while other points in or on the rigid body move in circular paths concentric to the axis and in planes perpendicular to the axis If the angular position θ of any given radius OX of the body is specified as a function of time then the position, velocity and acceleration of every point of the body are known (see Fig 5)

As an example, if two successive positions of OX are taken over a small time interval Δt then the instantaneous angular velocity ω of the rigid body at time t (see Fig 5b) is specified as

$$\omega = \frac{\text{change in angular displacement of } OX}{\text{time interval}} = \lim_{\Delta t\to 0}\frac{\Delta\theta}{\Delta t} = \frac{d\theta}{dt} \qquad (6)$$

Similarly, the instantaneous angular acceleration at time t is given by

$$\alpha = \frac{\text{change in angular velocity}}{\text{time interval}} = \lim_{\Delta t\to 0}\frac{\Delta\omega}{\Delta t} = \frac{d\omega}{dt} = \frac{d^2\theta}{dt^2} \qquad (7)$$

As is the case for linear velocity and accelera tion, both angular velocity and acceleration are vectors The vectors representing these quantities are drawn along the axis of rotation in a direction which represents the direction of advance of a right handed screw which is rotated in a manner designated by the angular velocity and acceleration respectively (see Fig 6)

If now a plane body such as a disk or wheel is pictured rotating about a fixed axis, the tangential velocity of any point B on this body resulting from the rotation is given by the vector product of the

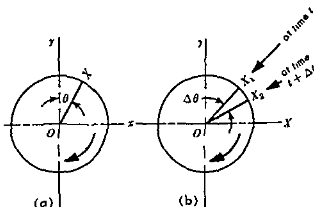


Fig 5 Wheel or disk rotating about a perpendicular axis through O, showing (a) angular quantities as func tion of time, (b) instantaneous angular velocity

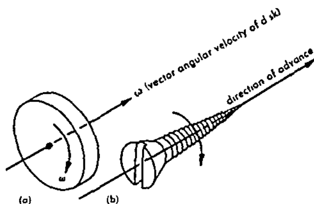


Fig 6 Vector representation of the angular velocity of (a) a rotating disk, (b) a rotating screw

$$W_i = \int_1^2 \mathbf{F}_i \cdot d\mathbf{r}_i \quad (24)$$

ere F_i is the resultant vector force acting on the rticle and dr_i is the vector displacement of the rticle as it travels from A to a neighboring point in time dt (see Fig 9)

For a rigid body considered as a system of such emental particles, the work done by all of the ternal forces acting on the rigid body (the inter al forces in a rigid body do no work) from Eqs 24) and (11) is

$$W_{1,2} = \int_1^2 \sum (\mathbf{F}_i \cdot d\mathbf{r}_i) = \int_1^2 \sum \left(m_i \frac{d^2\mathbf{r}_i}{dt^2} \right) \cdot d\mathbf{r}_i$$

$$= \int_{t_1}^{t_2} \sum m_i \frac{d}{dt}\left(\frac{d\mathbf{r}_i}{dt} \cdot \frac{d\mathbf{r}_i}{dt} \right) dt$$

$$= \int_{t_1}^{t_2} \frac{d}{dt} \sum \frac{m_i v_i^2}{2} dt = \int_1^2 d(T) = T_2 - T_1 \quad (25)$$

Thus, the change in the kinetic energy T of a rigid body in going from locations 1 to 2 in space is equal to the work done on the rigid body during this period Or, the rate of change of kinetic energy of the body is equal to the rate at which work is done by all external forces acting on the body If the forces acting are a function of the coordinates only and the work done is therefore independent of the path the body follows (gravitational forces, for example), the system is said to be conservative (frictionless without energy dissipation) Then the work done on the body by the forces acting on it, when this body moves in a conservative force field from locations 1 to 2 in space, is called the potential energy of the body at 2 with respect to 1, or is equal to the negative of the change in potential energy of the body in passage from 1 to 2. Then

$$W_{1,2} = -\Delta V_{1,2} \quad (26)$$

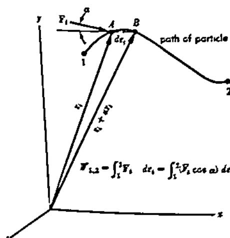


Fig. 9 Illustration of the work done by a force moving a mass particle along an arbitrary path

where $\Delta V_{1,2}$ is the change in potential energy of the body From Eqs (25) and (26)

$$T_1 + V_1 = T_2 + V_2 \quad (27)$$

Equation (27) is called the conservation of energy equation for a conservative system wherein the sum of kinetic and potential energies remains constant throughout the motion *See* CONSERVATION OF ENERGY, ENERGY, WORK

Rolling motion. The motion of a cylinder rolling without slipping on a surface offers an interesting and important application of Eqs (4) and (24) If a moving coordinate system is located at the center of mass, as shown in Fig 10, then $\omega_x = \omega_y = 0$ and $\omega_z = \omega$ Then the equations for linear and angular motion become

$$\Sigma F_x = ma_x = m\frac{dv_x}{dt} = m\frac{d^2x_o}{dt^2} \quad (28a)$$

$$\Sigma F_y = 0 \quad (28b)$$

$$M_o = I_o\frac{d\omega}{dt} \quad (28c)$$

The energy equation for such a conservative system is

$$\Delta T = -\Delta V \quad (29)$$

where

$$T = \tfrac{1}{2} mv_x^2 + \tfrac{1}{2} I_o\omega^2 \quad (30)$$

and V is the potential energy of the cylinder

The specification of no slipping requires, first, that the cylinder complete one revolution as the center O advances a distance of one circumference, or

$$r\theta = x \qquad r\left(\frac{d\theta}{dt}\right) = r\omega = \frac{dx}{dt} = v_x \quad (31)$$

and, second, that a frictional force f, acting at the point of contact is sufficiently large to create a moment M_o that satisfies Eq (28c)

If the cylinder in Fig. 10 starting from rest, rolls without slipping down the incline under the influence of gravity the motion can be determined from Eqs (28) which become for this case

$$W \sin\alpha - f = m\frac{d^2x_o}{dt^2} = m\frac{dv_x}{dt} \quad (32)$$

$$W \cos\alpha - N = 0 \quad (33)$$

$$rf = I_o\frac{d\omega}{dt} = (mK^2)\frac{d\omega}{dt} \quad (34)$$

where N is the normal force and K is the radius of gyration such that $mK^2 = I_o$ (*see* RADIUS OF GYRATION) Substitution of Eq. (34) into (32) gives

$$mg\sin\alpha - \frac{mK^2}{r}\frac{d\omega}{dt} = m\frac{dv_x}{dt} \quad (35)$$

and Eq (31) into (35) results in

$$\frac{dv_x}{dt} = \frac{g\sin\alpha}{1 + (K^2/r^2)} = a_x \quad (36)$$

tion can be written about the moving center of mass fixed within the projectile, instead of about some fixed point located exterior to the projectile, results in great simplification in the analysis of the motion *See* BALLISTICS, EXTERIOR

Moments and products of inertia To investigate the rotation of a rigid body, Eq (14) or (15) is used, wherein the resultant moment and the angular momentum are taken either about a point fixed in an inertial system or about the center of mass of the body In order to handle the equations most easily, however, it is usually essential that the point about which the moments and angular momenta are calculated be fixed in the body itself Therefore, if the body contains a point fixed in an inertial system, this point is taken as the origin of a coordinate system If no point within the body is fixed, then the center of mass is taken as the origin

Because the same equation, (14) or (15), describes the rotation of the body in either of these two cases, they are treated together Consider a set of axes xyz attached to the rigid body with either a fixed point or center of mass as origin O The elements or particles of the rigid body are at rest with respect to these axes, and as a result the velocity of any element dV relative to the origin O (Fig 8) is given by

$$\frac{d\mathbf{r}}{dt} = \mathbf{v} = \boldsymbol{\omega} \times \mathbf{r}$$

Then the angular momentum of the body about O, Eq (14*a*) becomes

$$\mathbf{H}_o = \int (\mathbf{r} \times \mathbf{v})\rho\, dV = \int \mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r})\rho\, dV \qquad (16)$$

Now $\mathbf{r}$ and $\boldsymbol{\omega}$ are vectors which can be expressed in terms of their components along x, y, z axes

$$\begin{aligned} \mathbf{r} &= \mathbf{i}x + \mathbf{j}y + \mathbf{k}z \\ \boldsymbol{\omega} &= \mathbf{i}\omega_x + \mathbf{j}\omega_y + \mathbf{k}\omega_z \end{aligned} \qquad (17)$$

Substituting Eqs (17) into (16) gives

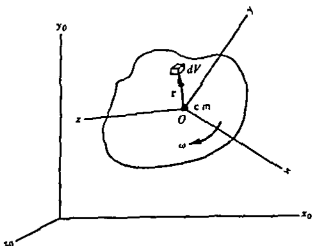


Fig 8 Rigid body of arbitrary shape with an xyz coordinate system attached at the center of mass and moving with the body

$$\begin{aligned} \mathbf{H}_o = {} & \mathbf{i}\int[\omega_x(y^2 + z^2) - \omega_y xy - \omega_z xz]\rho\, dV \\ & + \mathbf{j}\int[-\omega_x xy + \omega_y(x^2 + z^2) - \omega_z yz]\rho\, dV \\ & + \mathbf{k}\int[-\omega_x xz - \omega_y yz + \omega_z(x^2 + y^2)]\rho\, dV \end{aligned} \qquad (18)$$

The values ω_x, ω_y, ω_z are the same for all volume elements in the rigid body, and Eq (18) can be rewritten in a particularly useful way

$$\begin{aligned} \mathbf{H}_o = {} & \mathbf{i}(I_x\omega_x - I_{xy}\omega_y - I_{xz}\omega_z) \\ & + \mathbf{j}(-I_{yx}\omega_x + I_y\omega_y - I_{yz}\omega_z) \\ & + \mathbf{k}(-I_{zx}\omega_x - I_{zy}\omega_y + I_z\omega_z) \end{aligned} \qquad (19)$$

where

$$I_x = \int(y^2 + z^2)\rho\, dV \qquad I_y = \int(z^2 + x^2)\rho\, dV \qquad I_z = \int(x^2 + y^2)\rho\, dV \qquad (20)$$

are called the moments of inertia of the rigid body about the x, y, z axis through O, and

$$\begin{aligned} I_{xy} &= I_{yx} = \int xy\rho\, dV \\ I_{zx} &= I_{xz} = \int xz\rho\, dV \\ I_{yz} &= I_{zy} = \int yz\rho\, dV \end{aligned} \qquad (21)$$

are called the products of inertia These six quantities are geometrical constants associated with the body and do not vary with its motion or with time

If Eq (19) is substituted into Eq (14) or (15) and the time derivatives of the moving unit vectors $\mathbf{i}$, $\mathbf{j}$ $\mathbf{k}$ taken, the final equation becomes

$$\begin{aligned} \mathbf{M}_o = {} & \mathbf{i}(\dot{H}_x + \omega_y H_z - \omega_z H_y) \\ & + \mathbf{j}(\dot{H}_y + \omega_z H_x - \omega_x H_z) \\ & + \mathbf{k}(\dot{H}_z + \omega_x H_y - \omega_y H_x) \end{aligned} \qquad (22)$$

where H_x, H_y and H_z are the x y, and z components of $\mathbf{H}_o$ in Eq (19) *See* MOMENT OF INERTIA

Principal axes An important theorem exists which states that at any point in a rigid body it is possible to find and construct a set of rectangular coordinate axes such that the products of inertia vanish and only the three moments of inertia exist This is the principal axis theorem and such coordinates are called principal axes In particular in any rigid body that has two perpendicular planes of symmetry the coordinate planes of the principal axes coincide with the planes of symmetry

If the principal axes are chosen for the x y z coordinate system Eq (22) simplifies as follows

$$\begin{aligned} \mathbf{M}_o = {} & \mathbf{i}(I_x\dot{\omega}_x + (I_z - I_y)\omega_y\omega_z) \\ & + \mathbf{j}(I_y\dot{\omega}_y + (I_x - I_z)\omega_z\omega_x) \\ & + \mathbf{k}(I_z\dot{\omega}_z + (I_y - I_x)\omega_x\omega_y) \end{aligned} \qquad (23)$$

Equation (23) is called Euler s equation after the famous mathematician of the eighteenth century and can be used to solve the majority of rigid body dynamics problems *See* EULER S EQUATIONS OF MOTION

Work and energy relations The work done by the forces acting on a single mass particle (labeled i) as the particle moves from point 1 to point 2 in space is defined by the following vector equation

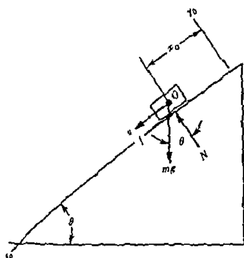


Fig 12 Block sliding down on inclined plane

Sliding motion Newton's second law, the principles of friction and experimental values of μ are needed to determine the motion of sliding rigid bodies.

$$\Sigma F = m \frac{dv_c}{dt}$$

and the two scalar equations of motion for this case are

$$mg \sin\theta - f = \frac{dv}{dt} = m\frac{d^2x}{dt^2} \qquad (43)$$

$$mg \cos\theta - N = 0 \qquad (44)$$

where

$$f/N = \mu' \qquad (45)$$

sults

$$v = g(\sin\theta - \cos\theta)t + v_0$$
$$x = g(\sin\theta - \cos\theta)t^2/2 + v_0 t$$
$$f = \mu\,(mg \cos\theta)$$

Combined rolling and sliding The problem of

... arge, the slope of the incline large, or the cylinder radius small the case easily arises where (Fig. 10)

$$\frac{f}{N} = \mu' < \frac{K^2 \tan\alpha}{K^2 + r^2} \qquad (46)$$

The cylinder then rolls and slides in a manner governed by the principle of kinetic friction and if μ is known as an experimental coefficient then the simultaneous solution of Eqs (32-34) and (46) gives the equations of motion See POINSOT'S METHOD [R F BO]

Bibliography L. Page, *Introduction to Theoretical Physics*, 3d ed, 1952, F W Sears *Mechanics, Heat and Sound*, 2d ed 1950, J L Synge and B A Griffith, *Principles of Mechanics*, 3d ed, 1959

Ring

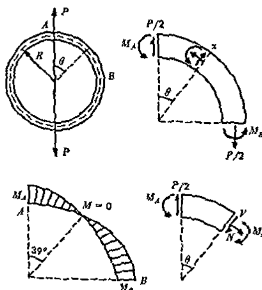
A tie member or chain link Tension or compression applied through the center of a ring produces bending moment, shear, and normal force on radial sections Because shear stress is zero at the boundaries of the section where bending stress is maximum it is usually neglected A quadrant of the ring is a curved bar with moments M_A and M_B at the ends as illustrated

An approximate solution for moments at the ends of the quadrant neglects curvature and employs the procedures used for statically indeterminate straight beams The numerically larger moment occurs at the load section A where $M_A = 0\,318PR$ and $M_B = 0\,181PR$ where P is load and R is radius The variation of moment is found by statics in terms of angle θ defining the section For tensile load $M_x = PR(0\,318 - \tfrac{1}{2} \sin\theta)$ Moment at A tends to increase the curvature, the ring becomes flatter at B The end moments have opposite signs and $M = 0$ near $\theta = 39°$

Important stresses occur at the inside and outside of sections where the normal and bending stresses reach maximum values At any section

$$S = S_b + \frac{P \sin\theta}{2A}$$

where S_b is found by curved bar theory and A is sectional area Or stresses can be calculated from the corrected straight beam formula



Stresses in rings

where g is the acceleration of gravity. The cylinder therefore rolls with a constant acceleration a_x.

The motion can also be determined by use of Eq. (29), the conservation of energy equation.

If the frictional force f were zero, then ω would be zero (see Eq. 34) and pure translation would take place at an acceleration of $g \sin \alpha$. The value of acceleration is therefore always less for the frictional (no slip) case but approaches the $g \sin \alpha$ value when the mass of the cylinder is so concentrated at the center that $K \to 0$. If the cylinder is a thin walled tube such that $K \to r$, then

$$\frac{dv_x}{dt} = \tfrac{1}{2} g \sin \alpha$$

If the cylinder is solid and uniform, $K^2 = r^2/2$, and for this case

$$\frac{dv_x}{dt} = \tfrac{2}{3} g \sin \alpha$$

The frictional force required to prevent slipping is

$$f = \frac{mgK^2 \sin \alpha}{r^2 + K^2} \tag{37}$$

From Eq. (33), the normal force is

$$N = mg \cos \alpha \tag{38}$$

Since $f = \mu N$ where μ is defined as the coefficient of static friction, Eqs. (37) and (38) show that in order for rolling to take place without slipping the coefficient must be such that

$$\mu \geq \frac{K^2 \tan \alpha}{K^2 + r^2} \tag{39}$$

See Friction, Statics

Instantaneous axis Consider a rigid body moving in a completely general manner. A point A is chosen on or within the body as an arbitrary base point and the velocity of A denoted by $\mathbf{v}_A$. The velocity of any other point B on or within the body is given by

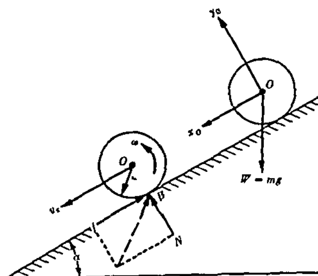


Fig. 10 Cylinder rolling down an inclined plane without slipping

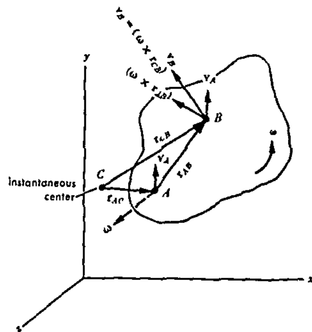


Fig. 11 Sketch showing location of instantaneous center of a rigid body moving in an arbitrary manner

$$\mathbf{v}_B = \mathbf{v}_A + (\boldsymbol{\omega} \times \mathbf{r}_{AB}) \tag{40}$$

where $\mathbf{r}_{AB}$ is the vector distance between A and B directed toward B. If the base point A is changed, the translational velocity $\mathbf{v}_A$ will be different but the angular velocity $\boldsymbol{\omega}$ will remain the same. The vector $\boldsymbol{\omega}$ pertains to the angular motion of the body as a whole and is to be regarded as a free vector since it does not depend upon a choice of base point.

Now, it is always possible to find a point C in space (not necessarily on or in the body) and an axis through this point parallel to $\boldsymbol{\omega}$ such that the velocity of any point B on or in the body is given at any instant by

$$\mathbf{v}_B = \boldsymbol{\omega} \times \mathbf{r}_{CB} \tag{41}$$

Therefore, it will also be true from Eq. (40) that

$$\mathbf{v}_A = \boldsymbol{\omega} \times \mathbf{r}_{CA}$$

Such a point C is called an instantaneous center (with an instantaneous velocity of zero) and the axis through C an instantaneous axis. The motion of a rigid body may therefore be considered as a succession of pure rotations about the instantaneous axis.

The instantaneous axis can always be found as follows. If $\mathbf{v}_A$ is the known velocity of a base point A (Fig. 11), then the instantaneous axis is found by shifting the axis of rotation to pass through a new point C located a vector distance $\mathbf{r}_{AC}$ from A such that

$$-\mathbf{v}_A = \boldsymbol{\omega} \times \mathbf{r}_{AC} \tag{42}$$

For the rolling cylinder case in Fig. 10 the instantaneous center (Eq. 42) is the point of contact B because

$$v_x = r_{BO}\omega$$

er the concept of a subfield of a field is defined

Ideals difference rings, homomorphism In elementary number theory it is often important to separate the ring I of integers into subsets defined by a divisibility condition For example there are the sets of even and odd integers. More generally, if m is a positive integer then the set of integers decomposes into m subsets I_0 I_1, , I_{m-1} where I_j is the set of integers which gives the remainder j on division by m Two integers a and b are in the same I_j if and only if $a - b$ is divisible by m If the condition holds one writes following Gauss

are the m sets I_0 I_1 , I_{m-1} The study of this ring gives a natural setting for important results of number theory

The construction just indicated can be carried in the

2 If m is in M and a is any element of R then am and ma are in M If $R = I$ the ring of integers then the set M of multiples of a fixed positive integer m is an ideal

As in this special case two elements a b of any ring R are said to be congruent modulo an ideal M if $a - b$ is in M The ring M can be decomposed into nonoverlapping congruence classes which of nce the nce na of

groups see GROUP THEORY) Also as in group theory the mapping which associates with every a of R the congruence class $[a]$ of R/M is the prime instance of a homomorphism If R and R' are any two rings a mapping $a \rightarrow f(a)$ of R into R' is a homomorphism if $f(a + b) = f(a) + f(b)$ and $f(ab) = f(a)f(b)$ If distinct el

the ring R/M where M is the ideal of elements mapped into 0 This basic result is the fundamental theorem of homomorphism for rings

Advanced aspects Ideals play an important role in higher arithmetic which deals with the arithmetic aspects of number fields Ideals are also basic in algebraic geometry In fact the theory of algebraic curves makes considerable use of a type of ideal theory called Dedekind ideal theory which includes the ideal theory of number fields On the other hand, higher-dimensional algebraic geometry can be founded on the ideal theory of so-called Noetherian rings

The structure theory of rings is essentially an analysis of the anatomy of rings The general idea is that of reducing the study of classes of rings to that of simpler classes Intertwined with this is the theory of representations of rings A representation of a ring is defined to be a homomorphism of a ring into a ring of endomorphisms of a commutative group *See* BOOLEAN ALGEBRA, LATTICE (MATHEMATICS), NUMBER THEORY [N J]

Bibliography E Artin et al *Rings with Minimum Condition* 1944, G Birkhoff and S MacLane *A Survey of Modern Algebra*, 1941, N Jacobson, *Lectures in Abstract Algebra*, vols 1 and 2, 1951 and 1953, N Jacobson *Structure of Rings*, 1956 D G Northcott *Ideal Theory*, 1953, P Samuel and O Zariski *Commutative Algebra*, 1958, B L. van der Waerden *Modern Algebra* vols 1 and 2, transl from 2d rev German ed, 1949–1950

Ripple tank

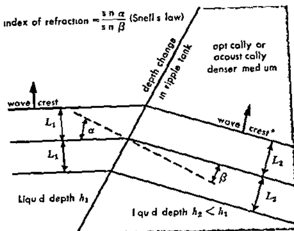
A shallow tray containing a liquid and equipped with a means for generating surface waves Ripple tanks are used to study a number of physical phenomena which can be described in terms of wave mechanics Water acoustic light and electromagnetic waves can be investigated with equal facility

Theory All wave motion can be described in terms of wave period T, or frequency $f = 1/T$, and wavelength L, hence celerity C or speed of propagation of a wave is defined as $C = fL$ The speed of surface waves on the liquid in the ripple tank is dependent upon density ρ, surface tension σ and depth of liquid h, thus

$$C^2 = \left(\frac{gL}{2\pi} + \frac{\sigma}{\rho}\frac{2\pi}{L}\right)\tanh\frac{2\pi h}{L}$$

n e ...d by the fundamental relation $L = C/f$ Therefore

$$C^2 = \left[\left(\frac{g}{2\pi f}\right)C + \left(\frac{2\pi f \sigma}{\rho}\right)\frac{1}{C}\right]\tanh\,(2\pi f)\,\frac{h}{C}$$



Wave refraction at a boundary between deep and shallow sections of a ripple tank

$$S = K\frac{M_c}{I} + \frac{P \sin\theta}{2A}$$

where K depends on end restraints, I is moment of inertia, and c is distance from centroids to extreme element The greatest stress (compressive for tensile load) occurs at the inside of the section where load is applied Where the greatest tensile stress occurs depends on the ring dimensions For a thick ring, whose radius of curvature is small compared to the dimensions, maximum tension may occur at top and bottom sections, whereas for thin rings with relatively large radius of curvature, greatest tension is at the side sections perpendicular to the load line Thin rings subjected to external hydrostatic pressure may fail by compressive buckling Deflection of the ring is most conveniently found by energy methods, which include the moments, shearing, and normal forces acting in the ring Bending has the greater influence in thin rings See CURVED BARS [W J KR]

Ring theory

a

isfying most but not all the properties of the addition and multiplication in the system of integers, 0 ±1, ±2 In precise terms a ring is a set R with two binary compositions called addition and multiplication whose results on an ordered pair (a,b), a,b in R, are denoted by $a+b$ and ab respectively These compositions must satisfy the following conditions

C $a + b$ and ab belong to R (closure)
A1 $a + b = b + a$ (commutative law)
A2 $(a + b) + c = a + (b + c)$ (associative law)
A3 There exists an element 0 (called zero) in R satisfying $a + 0 = a$ for every a in R
A4 For each a in R there exists an element $-a$ (called the negative of a) in R such that $a + (-a) = 0$
M1 $(ab)c = a(bc)$
D $a(b + c) = ab + ac$, $(b + c)a = ba + ca$ (distributive laws)

In the ring I of integers (addition and multiplication as usual) there are further conditions, for example, the commutative law of multiplication ($ab = ba$) and the cancellation law that if $a \neq 0$ and $ab = ac$ then $b = c$ See SET THEORY

Types of rings. The importance of the concept of a ring stems from the fact that it enbraces many special cases which are fundamental in all branches of mathematics Thus it includes the ring I of integers, the ring R_0 of rational numbers, the ring $R^\#$ of real numbers, the ring C of complex numbers, various rings of functions, rings of matrices, and so on An example of matrix rings is the following Let $R_n{}^\#$ denote the collection of all the n by n arrays or matrices

$$A = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \\ a_{n1} & a_{n2} & \cdot & a_{nn} \end{pmatrix}$$

whose entries a_{ij} are taken in the ring $R^\#$ of real numbers Two such matrices $A = (a_{ij})$ and $B = (b_{ij})$ are regarded as equal if and only if $a_{ij} = b_{ij}$ for every $i,j = 1, 2, \quad , n$ If $A = (a_{ij})$, $B = (b_{ij})$ then $A + B$ is defined to be $S = (s_{ij})$ where $s_{ij} = a_{ij} + b_{ij}$ and AB is defined to be $P = (p_{ij})$ where $p_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \quad + a_{in}b_{nj}$ The conditions C, A1–A4, M, and D are fulfilled, so that $R_n{}^\#$ is a ring If $n = 1$ this is essentially the same as the ring $R^\#$, but if $n > 1$, then $R_n{}^\#$ differs from $R^\#$ because the commutative law of multiplication does not hold, that is, there are instances in which $AB \neq BA$ The example $R_n{}^\#$ has a geometric counterpart, namely, the ring of linear transformations of an n dimensional vector space In fact these rings are isomorphic in the sense defined below The first example of a ring in which the commutative law of multiplication does not hold is the system Q of quaternions introduced by W R Hamilton as an appropriate tool for studying rotations in 3 dimensional space This system is the set of all the expressions of the form $a1 + bi + cj + dk$ where a, b, c, d are in $R^\#$ Addition is defined by adding corresponding coefficients as in ordinary algebra, and multiplication is defined to satisfy the associative and distributive laws and the following rules for the quaternion units 1 i, j, k $1u = u = u1$ for $u = 1$, i, j, k, $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$, $ki = -ik = j$ See QUATERNIONS

The conditions A1–A4 on the addition composition are exactly equivalent to the statement that any ring is a commutative group relative to its addition composition This group is called the additive group of the ring The algebraic system consisting of the set of elements of a ring together with its multiplication composition is called the multiplicative semigroup of the ring See GRAPH THEORY

Various classes of rings are singled out by imposing conditions on the multiplicative semigroup Thus integral domains are rings in which the product of nonzero elements is nonzero Division rings are rings whose sets of nonzero elements are groups relative to the multiplication composition, and

commutative

f quaternions

$R^\#$ and C are

fields Important instances of integral domains are the rings $F[x_1\ x_2, \quad x\ ]$ of all formal polynomials in the letters x_i with coefficients taken out of some field F

If R is a ring then a subring of R is a subset S of R which is a ring with respect to the addition and multiplication defined in R The conditions for this are that if s_1 and s_2 belong to S then so do $s_1 - s_2[= s_1 + (-s_2)]$ and s_1s_2 In a similar man

Level 3 can radiate directly to the ground state 0 emitting f_{30}, or it may first make a transition to level 2, which subsequently radiates to the ground state, etc Since the total energy emitted in these two alternative means of making transitions from

River

A water stream of natural origin which flows across the surface of a continent or island A river is part of a river system which drains a topographically related section of land surface known as a river basin The system begins in the precipitation which falls on a rock soil or vegetation covered surface and immediately becomes surface runoff or eventually appears as snow and ice meltwater or underground drainage Such a system may be divided into headwater streams tributary streams and the main stem. The headwaters are in springs marshes lakes or small upper streams generally in the highest relative elevation in a basin. A river ends in a mouth where it may discharge into a major lake a dry basin of interior drainage (playa) an inland sea the ocean

Terminology Like many words which have long been in general use, the term river is somewhat elastic in meaning In English usage the main stem of a stream system is nearly always designated as a river but so are all important tributaries, and even some secondary tributaries A tributary may also be known as a fork branch or creek and may have the same volume of flow as other streams called rivers Smaller headwater streams are usually creeks or brooks.

...these valley patterns *see* FLOOD PLAINS FLUVIAL EROSION LANDFORMS

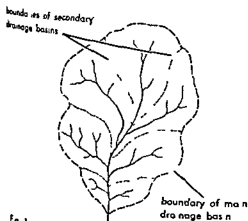


Fig 1 Maplike diagram of a drainage basin Note that such basins are composed of a system of secondary basins.



Fig 2 Cartographic diagram illustrating stream divide and basin patterns in a dendritic drainage system Streams shown by black lines divides by white

Rivers may be described by the pattern of the system of which they are part by their length velocity volume of discharge, and the nature of water flowing within them Most rivers are part of a dendritic drainage pattern but some responding to the underlying geologic structure are in radial annular rectangular or trellised (latticelike) pattern In some limestone regions a karst (enclosed depression) drainage may be found with underground rivers A few rivers such as the Nile in its lower reaches are exotic and flow for considerable distances without receiving drainage of any consequence from tributaries Such river reaches always occur in arid regions

River regime and flow patterns The regime is directly dependent on the climate of the region or regions involved It also is influenced by the size of the drainage basin funneling upon the stream the direction of flow the conditions of vegetative cover and the nature of the surface geology topography and soil conditions in the basin Few if any streams have completely stable conditions of flow the rule is variation from day to day season to season and year to year Study of these variations and their causes is an important part of the science of hydrology (*see* SURFACE WATER)

In arid regions intermittent streams are common The flow of an intermittent stream may *fluctuate* markedly from nothing to flood stage within a matter of minutes if a storm of sufficient extent and intensity covers part or all of its drainage area Normally dry channels of these streams are called arroyos or wadis *See* DESERT EROSION FEATURES

For a given liquid and frequency of ripple generation the parenthetical terms will be constants and the speed of propagation will depend only upon liquid depth h. However, as the depth becomes large, the hyperbolic tangent term rapidly approaches unity and the celerity equals a constant C_0 which also depends upon the frequency and fluid properties, hence

$$C_0^2 = \left(\frac{g}{2\pi f}\right) C_0 + \left(\frac{2\pi f \sigma}{\rho}\right) \frac{1}{C_0}$$

A graph of ratio C_0/C versus water depth may therefore be constructed to serve as the ripple tank calibration.

Analogy with wave phenomena. Diffractive phenomena associated with the propagation of sound, light, and electromagnetic waves can readily be studied in the ripple tank by observing that the ratio of celerities obtained above can be interpreted in the following manners: (1) for sound waves, C_0/C is the acoustic index of refraction, (2) for light waves C_0/C is the optical index of refraction, and (3) for electromagnetic waves, C_0/C is the square root of the dielectric constant (for those materials whose magnetic permeability is near unity). See REFRACTION OF WAVES.

Control of water depth in the ripple tank by contouring the bottom permits the simulation of a variable or discontinuous media for wave propagation. For example, a step change in the depth as shown in the diagram results in a wave diffraction in accord with Snell's law. Acoustical and optical wave diffractions have been studied in ripple tanks and phase fronts near two dimensional models of antenna structures and radomes have also been investigated. Ripple tank models of large scale water wave motions in harbors and along seacoasts can be useful provided that the surface tension term in the celerity equations is small compared with the gravitational term. Water depths of approximately 1 m and frequencies of 1 cps have been used successfully in this type of investigation.

Experimental equipment. The usual ripple tank consists of a glass or plastic plate with vertical sides to contain the liquid. Electronically driven probe vibrators are used to excite the water surface at a given frequency. Synchronously chopped light is directed through the tank to a ground glass screen on which the phase front shadow patterns appear stationary. Reflections from the walls of the tank are reduced to a minimum by placing folded wire screens and cloth around the boundaries. The head of a pin vibrating vertically can be made to serve as a radiating probe.

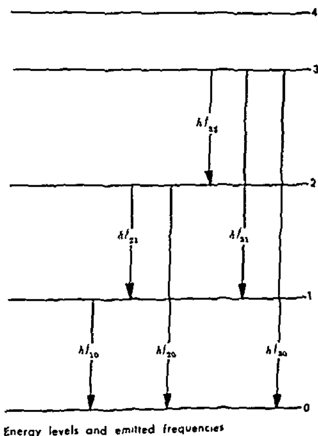
If care is taken to measure and adjust water levels accurately, depths as small as 0.1 mm can be used. At an exciting frequency of 20 cps a range of depths of 5–0.1 mm will result in a refractive index of 1–2.5 which would also correspond to a dielectric constant of approximately 1–6. At these small depths only short paths along which the waves travel can be used because of attenuation of the wave. See SHADOWGRAPH OF FLUID FLOW; WAVE MOTION IN LIQUIDS. [D.R.F.H.]

Ripple voltage

The total voltage across the load resistor of a rectifier minus the average voltage across the same resistor. The ripple can be expressed as a Fourier series. The fundamental frequency of the ripple voltage of single-phase half wave circuits is the same as that of the ac input. For a single-phase full wave circuit the fundamental frequency is twice that of the ac input voltage, while for a three-phase half wave, the fundamental frequency is three times that of the ac supply. To reduce the ripple voltage a low pass filter is usually placed between the rectifier and the load. The filter is more efficient in reducing the ripple voltage if the fundamental frequency of the ripple voltage is high. See RECTIFIER. [D.L.W.]

Ritz's combination principle

The empirical rule (W. Ritz, 1905) that sums and differences of the frequencies of spectral lines often equal other observed frequencies. The rule is an immediate consequence of the quantum mechanical formula $hf = E_i - E_f$ relating the energy hf of an emitted photon to the initial energy E_i and final energy E_f of the radiating system, h is Planck's constant and f is the frequency of the emitted light. For example, the figure shows the photon energies hf_{32}, hf_{31}, hf_{30} associated with transitions from level 3 to lower lying levels, etc.



Energy levels and emitted frequencies

level 3 can radiate directly to the ground state 0 emitting f_{30} or it may first make a transition to level 2, which subsequently radiates to the ground state, etc. Since the total energy emitted in these two alternative means of making transitions from 3 to 0 is exactly the same, namely $E_3 - E_0$, it follows that $hf_{30} = hf_{32} + hf_{20}$. Similarly, $hf_{30} = hf_{32} + hf_{21} + hf_{10}$ etc. *See* ATOMIC STRUCTURE AND SPECTRA, QUANTUM MECHANICS [E G]

River

A water stream of natural origin which flows across the surface of a continent or island. A river is part of a river system which drains a topographically related section of land surface known as a river basin. The system begins in the precipitation which falls on a rock, soil, or vegetation covered surface and immediately becomes surface runoff, or eventually appears as snow and ice meltwater or underground drainage. Such a system may be divided into headwater streams, tributary streams, and the main stem. The headwaters are in springs, marshes, lakes, or small upper streams generally in the highest relative elevation in a basin. A river ends in a mouth where it may discharge into a major lake, a dry basin of interior drainage (playa), an inland sea, or the ocean.

Terminology. Like many words which have long been in general use, the term "river" is somewhat elastic in meaning. In English usage the main stem of a stream system is nearly always designated as a

same volume of flow as other streams called rivers. Smaller headwater streams are usually creeks or brooks.

Rivers flow in channels or water courses and develop many distinctive valley features by erosion and deposition. For details of form and character of these valley patterns *see* FLOOD PLAINS, FLUVIAL EROSION LANDFORMS.

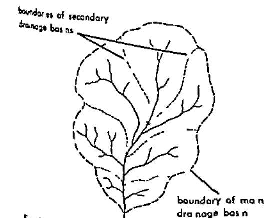


Fig. 1 Maplike diagram of a drainage basin. Note that such basins are composed of a system of secondary basins.



Fig. 2 Cartographic diagram illustrating stream divide and basin patterns in a dendritic drainage system. Streams shown by black lines, divides by white.

Rivers may be described by the pattern of the system of which they are part, by their length, velocity, volume of discharge, and the nature of water flowing within them. Most rivers are part of a dendritic drainage pattern, but some, responding to the underlying geologic structure, are in radial, annular, rectangular, or trellised (latticelike) pattern. In some limestone regions a karst (enclosed depression) drainage may be found, with underground rivers. A few rivers, such as the Nile in its lower reaches, are exotic, and flow for considerable distances without receiving drainage of any consequence from tributaries. Such river reaches always occur in arid regions.

River regime and flow patterns. The regime is directly dependent on the climate of the region or regions involved. It also is influenced by the size of the drainage basin funneling upon the stream, the direction of flow, the conditions of vegetative cover, and the nature of the surface geology, topography, and soil conditions in the basin. Few if any streams have completely stable conditions of flow; the rule is variation from day to day, season to season, and year to year. Study of these variations and their causes is an important part of the science of hydrology (*see* SURFACE WATER).

In arid regions intermittent streams are common. The flow of an intermittent stream may fluctuate markedly from nothing to flood stage within a matter of minutes if a storm of sufficient extent and intensity covers part or all of its drainage area. Normally dry channels of these streams are called arroyos or wadis. *See* DESERT EROSION FEATURES.

Under more humid climatic conditions the channels of streams of sufficient volume to be called rivers are only occasionally dry Fluctuations of flow nonetheless are found everywhere For example, natural flow near the mouth of the Tennessee has varied between 4500 and 500 000 ft³/sec In middle latitudes the season of low flow is generally summer when evaporation and transpiration within the basin are greatest High water may come during autumn winter or spring depending on temperature conditions and the time of heaviest precipitation Storage of large volumes of water such as snow over frozen ground characteristically causes early spring floods in the Great Plains region of the United States when a rapid thaw takes place

Within high latitude areas of the Northern Hemisphere high water inevitably occurs in spring on the northward flowing rivers because melting progresses from headwater to mouth and the flow of water released upstream is barred by ice dams remaining downstream The rivers of Siberia are notable examples of this condition with broad flooding over lowland plains

In low latitude areas on the other hand high water is directly related to seasonal maxima of rainfall but high altitude conditions may complicate the regime in most zones Where there is a pronounced dry season as on the Indian peninsula high water occurs soon after the onset of the rainy season when the moisture requirements of hitherto dormant vegetation are still low In all parts of the world altitudinal conditions may influence the regime of a stream in another manner Where headwaters are in extensive high mountain areas with heavy winter accumulations of snow and ice high water occurs at the season of heaviest melt early or mid summer The Columbia Ganges Indus and Rhine rivers all show this influence in their regimes.

Surface materials and the nature of vegetative cover also influence the regime The more continuous the forest or grass cover in general the more stable the volume of discharge Soil conditions which favor easy infiltration also promote more equitable flow as on the sands of the Atlantic and Gulf Coastal Plain of the United States

Water qualities Every river is an agent of erosion as well as an agent of drainage Many mineral materials other than water consequently are constantly in motion where a river flows These materials are transported by water in solution in suspension and as bed load For discussion of stream erosion transport deposition and associated land forms *see* STREAM TRANSPORT AND DEPOSITION

The high capacity of water as a solvent imparts many different qualities to river water as a solution The great majority of rivers are fresh water but a few are saline (relatively high salt content) All rivers however contain perceptible amounts of mineral material in water solution In most cases this is calcium the most common cause of "hard water but any of the elements soluble in water may be found such as magnesium potassium sodium silicon nitrogen and the elements which combine with them to form salts The content of salts in solution is highest in the rivers of regions under desert or semiarid climates but calcareous materials derived from limestone may yield hard water in humid regions

Like most other bodies of water on the earth s surface a river also is a medium for the support of life from bacteria and simple forms of plant life to fish and amphibian mammal and bird wildlife. This capacity is related not only to the capacity of water to carry nutrient minerals in solution but also dissolved gases and particularly oxygen

River management The characteristics of rivers have made them important to human society No other natural feature excepting the soil has been more closely tied to the past progress of civilization for the majority of human beings Means of counteracting the vagaries of flow have been an important part of civil engineering for centuries (see RIVER ENGINEERING) This has been true in part because of the attractiveness of flood plains to agricultural occupance and the consequent need to avoid natural flooding It has also followed from man s need for water storage in order to live through drought seasons In modern times the problem of river management or river control has become much more difficult because of the rapid increase of population its concentration in dense settlements the vastly increased disposal of wastes in rivers and the larger number of purposes that rivers must serve simultaneously The general objects of river management are the conservation of natural flow for release at the times needed by man the confinement of

Discharge basin area and length of some of the world s major rivers

	Average discharge ft³ sec	Basin area mi²	Length, miles
Amazon	7 200 000	2,7 2 000	3900
La Plata Paraná	2 800 000	1 198,000	2450
Congo	2,000 000	1 175 000	2900
Yangtze	7 0 000	750 000	3100
Ganges-Brahmaputra	707 000	93 000	1800
Mississippi Missouri	620 000	1 243 00	3892
Mekong	600 000	350 000	2600
Mackenzie	450 000	682,000	2525
Nile	420 000	1 293 000	4053
St Lawrence	400 000	565 000	2150
Volga	350 000	592,000	2375
Lena	325 000	1 169 000	2860
Yenisei	No data	1 000 000	3550
Ob	No data	1 000 000	2800
Danube	315 000	347 000	1725
Orinoco	No data	5 0 000	1600
Zambezi	No data	513 000	2200
Indus	300 000	372 000	1700
Amur	No data	87 000	2900
Niger	No data	584 000	2600
Columbia	235 000	258,200	1214
Huang	116 000	400 000	2 00
Yukon	No data	330 000	2100
Sao Francisco	No data	252,000	1811
Euphrates	No data	430 000	1 00
Murray Darling	No data	414 000	2314

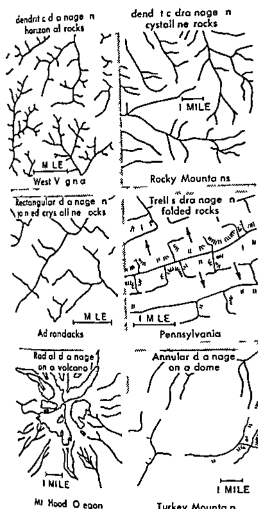


Fig 3 Cartog aph c d ag ams show ng types of stream patterns (After A K Lobeck Geomorphology McG aw-H ll 1939)

flood flow to the channel and planned areas of flood water storage and the maintenance of water quality at a level wh ch will yield the optimum benefit through multiple use The techniques of river man agement only recently have become well under stood the r practice is still very incomplete in part because the economics of river development is not well known. Domestic river development is an im portant responsibil ty of the United States Army Corps of Eng neers It also is the central responsi bil ty of the Tennessee Valley Authority and now an important objective of the Bureau of Reclama tion of the Department of the Interior

Of the major rivers in the world (see table) none is yet controlled or managed in the manner which modern eng neering administrative and biological techniques would permit. The closest approach to such management is made on some medium sized streams, the Tennessee, the Rhine and the Rhone for example. Some other rivers such as the San Joaquin in Cal fornia have been fully developed for a single purpose irrigation Commencing in the 1930s the greatest river regulation works of all his tory were undertaken The United States and the Soviet Union and since 1946 France have been

regulation of the Tennessee in the United States the Volga Don Canal the lower Volga dams and other works in the Soviet Union and the Rhone regulation in France

The greatest and potentially most productive works remain for the future These include plans for important work on the three largest streams of all the Amazon the La Plata Parana and the Congo These basins contain storage and power generation sites of several times the capacity of the largest hitherto developed Of the eight rivers hav ing basins of 1 000 000 or more square miles in ex tent only the Mississippi and Nile have more than minor control works Still other great streams offer ing major possibilities for physical development are the Yenisei Yangtze Huang Amur Mekong Chao

upon economic need availability of investment funds and political cooperation The need is patent for development of the Yangtze Huang Nile Ni ger Tigris Euphrates Danube Sao Francisco and lesser streams in densely settled underdeveloped areas It is therefore probable that the latter half of the twentieth century will be a period of extending control of these streams as political conditions permit [E A A]

Bibliography W H Hunter *Rivers and Estu aries* 1913 P H Kuenen *Realms of Water* 1956 F C Lane *Earth s Grandest Rivers* 1949 U S President s Water Resources Policy Commission *Ten Rivers in America s Future* vol 2 1950 *Large Rivers of the United States* USGS Circ 44 1949

River engineering

A branch of transportation engineering consisting of the physical measures which are taken to im prove the river and its banks

Most centers of civilization developed in the val leys of the world s rivers The people depended on alluvial plains for their agricultural economy and upon streams for domestic water and transporta tion Subsequently this reliance upon waterways has been expanded to include water for industrial as well as domestic consumption to provide eco nomical water power and to utilize the river for waste disposal With this expansion use of the streams for transportation has continued despite extensive developments of other transportation fa cilities Today the improvement of rivers for in land navigation is actively prosecuted in all parts of the civilized world Inland waterway traffic within the United States increased from 262 000 000 tons in 1947 to 384 000 000 tons in 1956



Fig 1 Aerial v ew show ng head of navigat on proj ect and uncontrolled and stab l zed sect ons of M s sour R ver near S oux C ty Iowa (U S Army Corps of Eng neers)

The measures that are taken to improve the river and its banks may include contraction of the river channel to improve navigation depths bank stabilization to minimize erosion which would destroy farm land cities and bridges creation of slack water pools by means of locks and dams or combinations of these means It may also include improvement of the channels to assist them in carrying flood flows and regulation of the rivers flows by upstream reservoir storage In approaching the problems of river engineering consideration must be given to the characteristics of the stream its slope meandering sediment load flow variations and other factors

River characteristics A stream is said to be in regimen if the major dimensions of the channel remain relatively constant and if it is neither aggrading (raising of the bed) nor degrading (lowering of the bed) The channel need not be fixed in position however many streams in regimen are constantly shifting their channels by eroding the banks at one location while building them at another

Most natural streams are in regimen with channel dimensions that are more or less unique to that stream This implies a balance between the energy forces of the flow the forces required to erode the bed and banks the sediment load and possibly other factors There is no universally accepted theory relating these factors but in general a stream in erodible alluvium will be wide and shallow if the banks are readily erodible or narrow and deep if the banks are erosion resistant

Channels may be generally classified as straight meandering (following an alignment consisting principally of pronounced bends) or braided (a number of interconnected channels presenting the appearance of a braid) These forms are influenced by many factors including the stream discharge the nature of the soils and the sediment load however they may be best correlated with the slope of the valley in which they are located Straight chan



Fig 2 Uncontrolled river



Fig 3 Pile dike contraction works

nels are found in valleys of either flat or steep slope, meandering channels occupy valleys of intermediate slope, and braided channels of streams in regimen occur on steep slopes Braided channels of streams not in regimen may occur on either flat, intermediate or steep slopes

Technical knowledge is inadequate to explain fully the relationship between stream form and valley slope, but it is necessary in river engineering to recognize it. For example many attempts to improve meandering channels by excavating straight channels have failed because the stream immediately began to erode its banks to reassert the meanders, meanwhile dumping excessive quantities of sediment into the channel downstream In like manner attempts to impose bends or curves on a channel in steep slopes must be considered with due caution.

River channel improvements These may consist of revetting the banks to prevent erosion and shifting of the channel realignment of the channel to provide smoother bends and a more regular alignment, contraction of an existent channel (particularly the contraction of a braided low water reach to provide a single effective low water channel) the provision of slack water pools by the construction of low dams and ship locks or the complete excavation of a new channel Problems in ... ing revetment, realignment or contraction occur most frequently in meandering streams in erodible alluvium

Contraction works Used primarily to confine the low and moderate flows of a wide shallow or braided stream to a single effective channel contraction works are required predominantly in meandering streams in erodible alluvium It is important that the rectified channel be planned with due regard to maintenance of regimen Alignment of the channel should be generally similar to that of the existent stream following a series of smooth bends rather than straight lines and maintaining essentially the same channel slope

Structures used in contraction consist of revetment in the concave portions of bends and of guide structures The latter are normally of pile dike or other permeable fence type construction designed to utilize the sedimentary and erosive characteristics of the stream in the initial shaping of the channel (*see* STREAM TRANSPORT AND DEPOSITION)

Where the position of the rectified channel deviates materially from that of the original channel the concave banks of bends may be excavated and revetted in the dry a pilot channel excavated and the stream encouraged to scour the channel to the desired dimensions In other cases the guide structures are constructed in stages contracting the channel and causing the opposing bank to erode progressively to the desired location The permeable guide structures serve to turn the current as desired yet permit deposition of sediments to build up the abandoned area behind them *See* REVETMENT

Locks and dams In some streams navigable depths are secured by relatively low head dams which create a series of slack water pools Locks consisting of gated chambers are provided to pass l

... e ga es a

with water drawn from the upper reservoir When the chamber has been filled to the level of the upper pool the upper gates are opened and the vessel passes through into the upper pool The reverse process is followed in going from the upper pool to the lower pool

The lift of the lock may vary from a few feet to over 100 ft The locks may be supplemented by gates through the dam These gates may be lowered



Fig 4 Lock and dam at Minneiska Minnesota

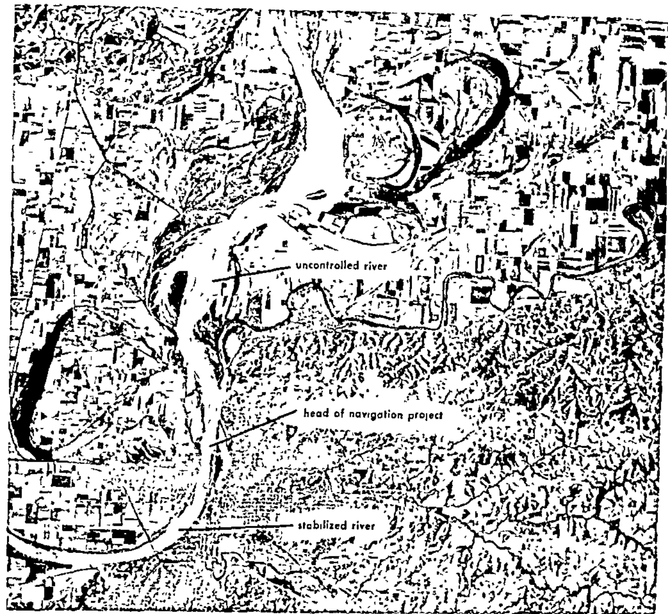


Fig. 1 Aerial view showing head of navigation project and uncontrolled and stabilized sections of Missouri River near Sioux City, Iowa (U S. Army Corps of Engineers)

The measures that are taken to improve the river and its banks may include contraction of the river channel to improve navigation depths, bank stabilization to minimize erosion which would destroy farm land, cities, and bridges; creation of slackwater pools by means of locks and dams; or combinations of these means. It may also include improvement of the channels to assist them in carrying flood flows and regulation of the rivers' flows by upstream reservoir storage In approaching the problems of river engineering, consideration must be given to the characteristics of the stream, its slope, meandering, sediment load, flow variations, and other factors.

River characteristics. A stream is said to be in regimen if the major dimensions of the channel remain relatively constant and if it is neither aggrading (raising of the bed) nor degrading (lowering of the bed). The channel need not be fixed in position, however; many streams in regimen are constantly shifting their channels by eroding the banks at one location while building them at another

Most natural streams are in regimen, with channel dimensions that are more or less unique to that stream This implies a balance between the energy [illegible] lating these factors, but, in general, a stream in erodible alluvium will be wide and shallow if the banks are readily erodible or narrow and deep if the banks are erosion resistant.

Channels may be generally classified as straight, meandering (following an alignment consisting principally of pronounced bends), or braided (a number of interconnected channels presenting the appearance of a braid). These forms are influenced by many factors, including the stream discharge, the nature of the soils, and the sediment load, however, they may be best correlated with the slope of the valley in which they are located. Straight chan-

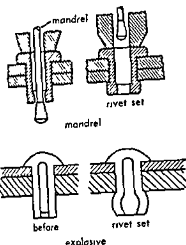


Fig. 2. Two types of blind rivet

holes are drilled and reamed to size and the rivet is driven to fill the hole completely Structural rivet ing uses punched holes

Small rivets (7/16 in and under) are used for general purpose work with head forms as follows flat, countersunk, button, pan, and truss (shown in Fig. 1) These rivets are commonly made of rivet steel although aluminum and copper are used for some applications The fillet under the head may be up to 1/32 in in radius

Large rivets (½ in and over) are used for structural work and in boiler and ship construction with heads as follows round top countersunk, button (most common), high button or acorn, pan, cone (truncated), and flat top countersunk

Boiler rivets have heads similar to large rivets with steeple (conical) added but have different proportions from large rivet heads in some cases

Special purpose rivets are tinners' rivets which have flat heads for use in sheet metal work, coopers rivets, that are used for riveting hoops for barrels, casks, and kegs, and belt rivets used for joining belt ends

Bl

In the explosive type, an explosive charge in the point is set off by a special hot iron, the explosion expands the point and sets the rivet

Standard material for rivets is open hearth steel (containing Mn P, S) with tensile strength 45 000-55 000 psi. Standards include acceptance tests for cold and hot ductility and hardness Materials for some special purpose rivets are aluminum and copper

[P H B]

Bibliography American Standards Association American Standard for Large Rivets (½" Nominal Diameter and Larger), B18 4-1950 (R1957) ASA, *American Standard for Small Solid Rivets*, B18.1 1955 T Baumeister (ed), *Marks' Mechanical Engineers' Handbook*, 6th ed, 1958, Industrial Fasteners Institute, *Fasteners Data Book*, 1950, V H Laughner and A D Hargan, *Handbook of Fastening and Joining of Metal Parts*, 1956

Riveted joint

The permanent joining of two or more machine or structural members, usually plates, by means of rivets The plates may be lapped or butted In the butt joint one or more cover plates must be used to accomplish the joining One of these cover plates is often made wider than the other (Fig 1) The rivets in the joint may be disposed in several ways to form single or multiple rows in a regular or staggered arrangement

Termin

terms of

between adjacent rivets

different pitches are used on adjacent gage lines

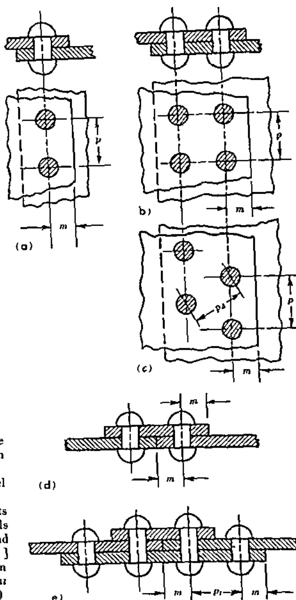


Fig 1 Typical riveted joints. (a, b, c) Lapped joints (d e) Butt joints

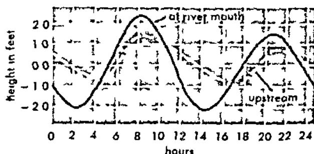
to the stream bed to permit free navigation during periods of adequate flow *See* DAM

Canals Canals are constructed to provide con nections between waterways or to bypass critical river reaches They may range from essentially open waterways, such as the Suez Canal, to complex systems of excavated waterways, dams, and high lift locks such as the Panama Canal, or they may be included in a system with locks, dams, and open river navigation as in the St Lawrence River They may also include channels excavated through low-slope braided streams or swamp areas, as in the Illinois River *See* CANAL [DCB, WEJ]

Bibliography T Blench, *Regime Behavior of Canals and Rivers*, 1957, R S Rowe, *Bibliography of Rivers and Harbors and Related Fields in Hydraulic Engineering*, 1953

River tides

Tides that occur in rivers emptying directly into tidal seas These tides show three characteristic modifications of ocean tides (1) The speed at which the tide travels upstream depends on the depth of the channel, $v \approx \sqrt{gh}$, where v is the speed, g the acceleration of gravity, and h the chan nel depth (2) The further upstream, the longer the duration of the falling tide and the shorter the duration of the rising tide (3) The range of the tide decreases with distance upstream *See* TIDE



River tide curves.

In a river the difference between the depths of water at high and low tides may be relatively large leading to a marked difference between the speeds at which high and low tides move In the Hudson River the low tide (at lower high water) takes 10 min longer than the high tide (at higher high wa ter) to reach Tarrytown, 24 nautical miles from the mouth, whereas the low tide takes 60 min longer than the high tide to reach Albany, 125 nautical miles from the mouth

The difference in depth between various points on the river also partially explains the second mod ification, or duration of fall and rise In addition the river flow, which may fluctuate widely, helps a falling tide but hinders a rising tide, increasing the difference in duration At the mouth of the Hudson the average fall lasts 6 hours 22 min, whereas the average rise lasts 6 hours 3 min At Tarrytown the values are 6 hours 33 min and 5 hours 52 min, and at Albany 7 hours 21 min and 5 hours 4 min, respectively

The third modification or decrease in tidal r upstream may be accounted for by loss of en of the water through friction with the sides bottom of the channel At the mouth of the Hu the average tidal range is 4 4 ft, whereas at T 131 nautical miles upstream, the range is 3 Although friction always saps energy from tide, if the channel becomes constricted with short distance, the water may be forced in smaller space, thus producing a larger tidal ra For example, Bristol Channel in Great Britai 40 miles wide at the mouth, where the average t range is 20 ft Within 80 miles the channel rows to 5 miles, at the mouth of the Avon Ri where the tidal range is 33 ft *See* TIDAL BORE.

Tides penetrate upstream until they encount dam, rapids, or falls In the Amazon a 10-ft at the mouth is detectable 450 nautical miles stream *See* FALL LINE [B

Bibliography A T Doodson and H D W burg, *Admiralty Manual of Tides*, 1941, H La *Hydrodynamics*, 1945, H A Marmer, *Tidal Dat Planes*, rev ed, U S Coast and Geodetic Sur Spec Publ 135, 1951, H A Marmer, *The T* 1926, J Proudman, *Dynamical Oceanograp* 1953, A C Redfield, The analysis of tidal phen ena in narrow embayments, *Papers Phys Ocean Meteorol*, 11(4) 1–36, 1950, J J Stoker, *W Waves*, 1957

Rivet

A short rod with a head formed on one end A ri is inserted through aligned holes in two or m parts to be joined, then by pressing the protr ing end, a second head is formed to hold the pa together permanently The first head is called t manufactured head and the second one the poi In forming the point a hold on or dolly bar is us to back up the manufactured head and the rivet driven preferably by a machine riveter For hig grade work such as boiler joint riveting the ri

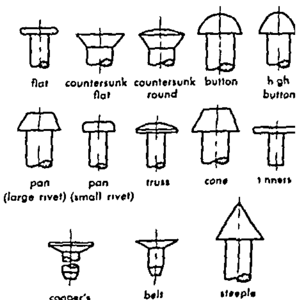


Fig 1 Standard rivet heads.

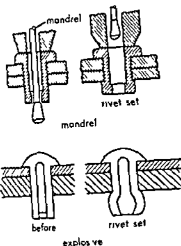


Fig 2 Two types of blind rivet

holes are drilled and reamed to size and the rivet is driven to fill the hole completely Structural riveting uses punched holes

Small rivets (7/16 in and under) are used for general purpose work with head forms as follows

some applications The fillet under the head may be up to 1/32 in in radius.

Large rivets (½ in and over) are used for structural work and in boiler and ship construction with heads as follows round top countersunk button (most common) high button or acorn, pan cone (truncated) and flat top countersunk

Boiler rivets have heads similar to large rivets with steeple (conical) added but have different proportions from large-rivet heads in some cases

Special purpose rivets are tinners' rivets which have flat heads for use in sheet metal work, cooper's rivets, that are used for riveting hoops for barrels casks, and kegs and belt rivets used for joining belt ends

In the explosive type, an explosive charge in the point is set off by a special hot iron the explosion expands the point and sets the rivet

Standard material for rivets is open hearth steel (containing Mn P, S) with tensile strength 45,000–55,000 psi. Standards include acceptance tests for cold and hot ductility and hardness Materials for some special purpose rivets are aluminum and copper

[P H B]

Bibliography American Standards Association American Standard for Large Rivets (½" Nominal Diameter and Larger) B18 4–1950 (R1957) ASA, American Standard for Small Solid Rivets B18.1 1955 T Baumeister (ed) Marks' Mechanical Engineers Handbook, 6th ed 1958 Industrial Fasteners Institute, Fasteners Data Book, 1950, V H Laughner and A D Hargan, Handbook of Fastening and Joining of Metal Parts 1956

Riveted joint

The permanent joining of two or more machine or structural members, usually plates, by means of rivets The plates may be lapped or butted In the butt joint one or more cover plates must be used to accomplish the joining One of these cover plates is often made wider than the other (Fig 1) The rivets in the joint may be disposed in several ways to form single or multiple rows in a regular or staggered arrangement

Terminology Riveted joints are described in terms of their dimensions Pitch p is the distance between adjacent rivets along the gage line Where different pitches are used on adjacent gage lines

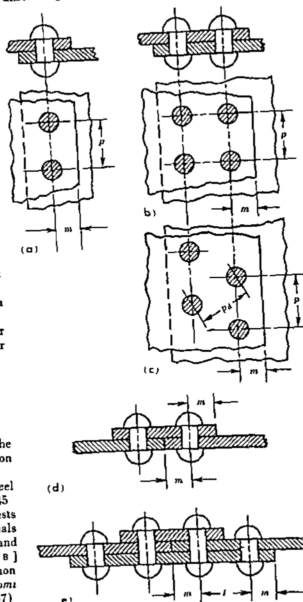


Fig 1 Typical riveted joints. (a b c) Lapped joints (d e) Butt joints

the largest is the pitch for the joint Gage line is the line through the centers of the rivets parallel to the edge of the plate A unit strip or length is equal in width to the pitch The distance p_t between two adjacent gage lines in the same plate is the back pitch or transverse pitch Diagonal pitch p_d is the distance between adjacent rivets on adjacent gage lines Margin m is the distance between a gage line and the edge of the plate The efficiency of a joint is the ratio of the weakest section of a unit strip to the strength of the same width of unperforated plate

Stresses. Lap joints are subject to eccentric loading, which brings about a bending of the joint (Fig 2) This in turn complicates the stress pattern on the various components in the joint, and the stresses calculated by straightforward assumptions of simple shear, bearing, or tensile loads must be increased by substantial factors to give adequate design stresses

A riveted joint in tension may fail in one of several ways (Fig 3) In butt joints with two or more plates and two or more rows of rivets the joint strength calculation is complicated by the uncertainty of the division of the load between the various rivets Although the strength of a joint is *usually considered* to be a function only of the strength of the rivets and plate, friction between the plates accounts for a large but indeterminate amount of load capacity



Fig 2 Eccentric loading bends a lap joint

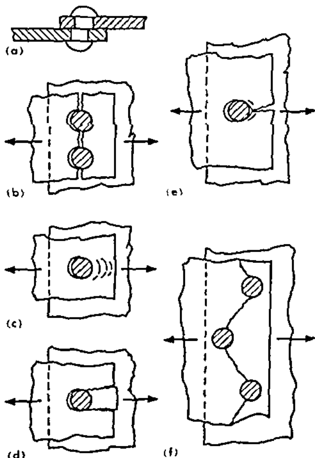


Fig 3 A riveted joint may fail from (a) shear, (b) rupture, (c) crushing of rivet or plate, (d) double shear through the margin, (e) rupture of the margin, or (f) zig zag tension

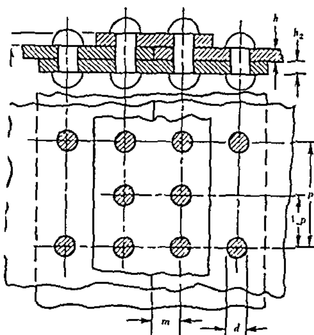


Fig 4 Dimensions of a riveted joint

Strength of riveted joint. The force F that a lapped riveted joint (Fig 4) can sustain depends on the stress S that can be withstood by the materials of which it is built and their dimensions (*see* STRENGTH OF MATERIALS) Thus the strength of the solid plate is $F = phS$ Tensile strength F_t at the outer gage line is $F_t = (p - d)hS$ Similarly, shear strength F_s of all rivets is

$$F_s = (2n_2 - n_1)\pi d^2 S_s/4$$

in which n_1 is the number of rivets in single shear n_2 is the number of rivets in double shear and d is the rivet diameter assuming all rivets to be the same size The crushing strength of rivets is

$$F_c = (n_2 h + n_1 h_2) d S_c$$

where h_2 is thickness of wider strap

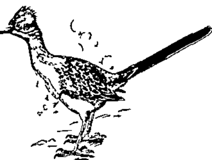
For maximum efficiency e of a joint $F_t = F_c = F_s$ Under this optimum condition

$$e_{max} = \frac{[n_2 + (n_1 h_2/h)]S_c}{[n_2 + (n_1 h_2/h)]S_c + S}$$

Resistance in shear does not appear in the equation for maximum efficiency *See* BOLTED JOINT JOINT (MECHANICAL), STRUCTURAL CONNECTIONS WELDED JOINT [L S L]

Road runner

A bird *Geococcyx californianus*, a member of the family Cuculidae the cuckoos The road runner is a rather odd looking chickenlike bird, with a long tail moderately long neck, and strong legs It is



ood runner Geococcyx cal forn anus length to 2 (From E L. Palmer F eldbook of Natural H story G aw H ll 1949)

ane ts enemies espe

a

orp ons and in ects They ometimes aes oy rd eggs and for this reason are considered harm ul, but such eggs are a very minor part of their

[J D B]

Robin

The mo t familiar of all American songbirds *Turdus m gratorius* a member of the thrush family The robin derives its name from a superficial re-



he rob n Turdus m grato us length to 10¾ n (From L Palmer F eldbook of Natu al H story McG aw-H ll 949

seml lance to tl e European rol in The rol in nests tl r gl o t N rth America from the tree lim t w tl ward except the southern portion of the c a tal plain In the fall it u ually migrates out of mu l f tl is territ ry and congregates f r the win ter in great r o ts in the southern Un ted States In tl e central states s me indivi l als may migrate, wh le thers seek shelter in the river valleys and reappear in towns and uplands early in the spring or during warm spells in winter *See* PASSERIFORMES [J D B]

Rochelle salt

Sodium potassium tartrate tetrahydrate



$$KNaC_4H_4O_6 \cdot 4H_2O$$

a crystalline solid known for its anomalously large dielectric constant and piezoelectric response for an electric field along the crystallographic *a* axis Because of its large piezoelectric effect Rochelle salt is widely used in microphones earphones, and phonograph pickup cartridges *See* PIEZOELECTRICITY

Between two critical temperatures (Curie points) −18°C and +24°C Rochelle salt is spontaneously electrically polarized along the *a* axis that is it has a permanent electric dipole moment A macroscopic Rochelle salt crystal consists in general of regions with opposite spontaneous polarization (domains) The spontaneous polarization can be aligned and reversed in the whole crystal by an electric field along the *a* axis or by a shear stress in the plane perpendicular to the *a* axis *See* FERROELECTRICS

Rochelle s [illegible] tremes of re [illegible] of crystallizat [illegible] humidity (25°C) and going into solution in water absorbed from ambient air at humidities above 85% It is also damaged by temperatures approaching 55°C (130°F) at which point it loses water of crystallization and breaks up into sodium and potassium tartrate [W K]

Bibliography W P Mason *Piezoelectric Crystals and Their Application to Ultrasonics* 1950

Rock

A relatively common aggregate of mineral grains Some rocks consist essentially of but one mineral species (monomineralic such as quartzite composed of quartz) others consist of two or more minerals (polymineralic such as granite composed of quartz, feldspar and biotite) Rock names are not given for those rare combinations of minerals that constitute ore deposits, such as quartz, pyrite and gold In the popular sense rock is considered also to denote a compact sub tance one with some c herence but geologically friable volcanic ash also is a rock A genetic classification of rocks is shown in the following list

- Igneous
  - Intrusive
    - Plutonic (deep)
    - Hypabyssal (shallow)
  - Extrusive
    - Flow
    - Pyrocla tic (explosive)
- Sed mentary
  - Cla tic (mechanical or detrital)
  - Chemical (crystall ne or precipitated)

the largest is the pitch for the joint Gage line is the line through the centers of the rivets parallel to the edge of the plate A unit strip or length is equal in width to the pitch The distance p_t between two adjacent gage lines in the same plate is the back pitch or transverse pitch Diagonal pitch p_d is the distance between adjacent rivets on adjacent gage lines Margin m is the distance between a gage line and the edge of the plate The efficiency of a joint is the ratio of the weakest section of a unit strip to the strength of the same width of unperforated plate

Stresses. Lap joints are subject to eccentric loading, which brings about a bending of the joint (Fig 2) This in turn complicates the stress pattern on the various components in the joint, and the stresses calculated by straightforward assumptions of simple shear, bearing, or tensile loads must be increased by substantial factors to give adequate design stresses

A riveted joint in tension may fail in one of several ways (Fig 3) In butt joints with two or more plates and two or more rows of rivets the joint strength calculation is complicated by the uncertainty of the division of the load between the various rivets Although the strength of a joint is usually considered to be a function only of the strength of the rivets and plate friction between the plates accounts for a large but indeterminate amount of load capacity

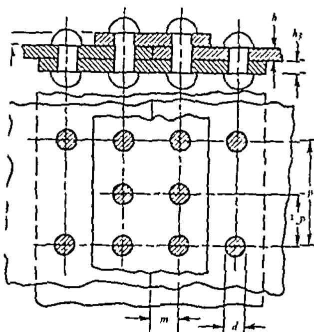


Fig 4 Dimensions of a riveted joint



Fig 2 Eccentric loading bends a lap joint

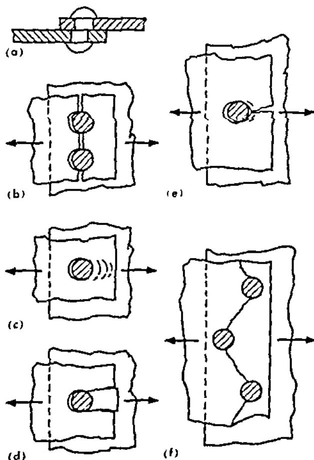


Fig 3 A riveted joint may fail from (a) shear, (b) rupture, (c) crushing of rivet or plate, (d) double shear through the margin, (e) rupture of the margin, or (f) zigzag tension

Strength of riveted joint. The force F that a lapped riveted joint (Fig 4) can sustain depends on the stress S that can be withstood by the materials of which it is built and their dimensions (*see* STRENGTH OF MATERIALS) Thus the strength of the solid plate is $F = phS$ Tensile strength F_t at the outer gage line is $F_t = (p - d)hS$ Similarly shear strength F_s of all rivets is

$$F_s = (2n_2 - n_1)\pi d^2 S_s/4$$

in which n_1 is the number of rivets in single shear n_2 is the number of rivets in double shear and d is the rivet diameter assuming all rivets to be the same size The crushing strength of rivets is

$$F_c = (n_2 h + n_1 h_2)dS_c$$

where h_2 is thickness of wider strap

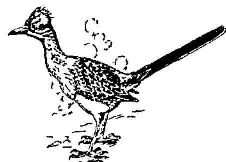
For maximum efficiency e of a joint $F_t = F_c = F_s$ Under this optimum condition

$$e_{max} = \frac{[n_2 + (n_1 h_2/h)]S_c}{[n_2 + (n_1 h_2/h)]S_c + S}$$

Resistance in shear does not appear in the equation for maximum efficiency *See* BOLTED JOINT JOINT (MECHANICAL), STRUCTURAL CONNECTIONS WELDED JOINT [L S L]

Road runner

A bird *Geococcyx californianus*, a member of the family Cuculidae, the cuckoos The road runner is a rather odd looking chickenlike bird, with a long tail, moderately long neck, and strong legs It is



The road runner, *Geococcyx californianus* length to 2 ft (From E L Palmer, Fieldbook of Natural History, McGraw Hill 1949)

terrestrial, and runs to escape its enemies espe cially man attaining speeds up to 18 mph

Road runners feed primarily upon snakes lizards scorpions and insects They sometimes destroy bird eggs and for this reason are considered harm ful, but such eggs are a very minor part of their d et ... the ... uth

[J D B]

Robin

The most familiar of all American songbirds *Tur dus migratorius* a member of the thrush family The robin derives its name from a superficial re-



The rob n *Turdus m gratorius* length to 10½ in (From E L Palmer F eldbook of Natural H story McGraw Hill 1949)

semblance to the European robin The robin nests throughout North America from the tree limit southward except the southern portion of the coastal plain In the fall it usually migrates out of much of this territory and congregates for the win ter in great roosts in the southern United States In the central states some individuals may migrate ... shelter in the river valleys and reappear in towns and uplands early in the spring, or during warm spells in winter *See* PASSERIFORMES [J D B]

Rochelle salt

Sodium potassium tartrate tetrahydrate,



$$KNaC_4H_4O_6 \cdot 4H_2O$$

a crystalline solid known for its anomalously large dielectric constant and piezoelectric response for an electric field along the crystallographic *a* axis Because of its large piezoelectric effect, Rochelle salt is widely used in microphones, earphones and phonograph pickup cartridges *See* PIEZOELECTRICITY

Between two critical temperatures (Curie points), −18°C and +24°C, Rochelle salt is spon taneously electrically polarized along the *a* axis, that is, it has a permanent electric dipole moment A macroscopic Rochelle salt crystal consists in gen eral of regions with opposite spontaneous polari zation (domains) The spontaneous polarization can be aligned and reversed in the whole crystal by an electric field along the *a* axis or by a shear stress in the plane perpendicular to the *a* axis *See* FERROELECTRICS

Rochelle salt is susceptible to damage by ex tremes of relative ambient humidity, losing water of crystallization from its surface in air below 35% humidity (25°C) and going into solution in water absorbed from ambient air at humidities above 85% It is also damaged by temperatures approach ing 55°C (130°F) at which point it loses water of crystallization and breaks up into sodium and potassium tartrate [W K]

Bibliography W P Mason, *Piezoelectric Crys tals and Their Application to Ultrasonics*, 1950

Rock

A relatively common aggregate of mineral grains Some rocks consist essentially of but one mineral species (monomineralic, such as quartzite, com posed of quartz), others consist of two or more minerals (polymineralic such as granite, composed of quartz feldspar, and biotite) Rock names are not given for those rare combinations of minerals

coherence, but geologically, friable volcanic ash also is a rock A genetic classification of rocks is shown in the following list

- Igneous
  - Intrusive
    - Plutonic (deep)
    - Hypabyssal (shallow)
  - Extrusive
    - Flow
    - Pyroclastic (explosive)
- Sedimentary
  - Clastic (mechanical or detrital)
  - Chemical (crystalline or precipitated)

Organic (biogenic)
Metamorphic
Cataclastic
Contact metamorphic and pyrometasomatic
Regional metamorphic (dynamothermal)
Hybrid
Metasomatic
Migmatitic

Exceptions to the requirement that rocks consist of minerals are obsidian, a volcanic rock consisting of glass, and coal, a sedimentary rock which is a mixture of organic compounds *See* COAL, VOLCANIC GLASS

Igneous rocks Igneous rocks are those that have solidified from a molten condition The parent material is magma—a natural, hot, mutual solution of silicates with minor amounts of water and other volatiles Igneous rocks are divided into those which crystallized before magma reached the earth's surface (intrusive rocks) and those that solidified at the surface, some as layers of lava (the extrusive flow rocks) and others as pyroclastic debris in explosive eruption (Table 1) *See* MAGMA, PYROCLASTIC ROCKS

Chief elements in igneous rocks are oxygen (O), silicon (Si), aluminum (Al), iron (Fe), magnesium (Mg), calcium (Ca) sodium (Na), and potassium (K) Compositions range from about 40% SiO_2 (peridotites) to as much as 70% SiO_2 (granites) Silica poor rocks contain relatively large amounts of Ca, Mg and Fe^2 (basic rocks) whereas silica rich types contain larger amounts of Na and K (acidic rocks) *See* IGNEOUS ROCKS

Sedimentary rocks Clastic sedimentary rocks (consisting of mechanically transported particles) are subdivided on the basis of particle size (Table 2) Those having intermediate and fine grain sizes are further subdivided on the basis of composition (Fig 1) Other significant clastic rocks are those consisting of detrital calcite the calcarenites and calcilutites *See* CALCARENITE

Textures of clastic rocks derive from grain size sorting, form and arrangement Form includes sphericity (shape) the degree to which a particle approximates a sphere, and roundness the measurement of the sharpness of edges and corners Only glacial sedimentary rocks (tillite) do not show layering or stratification

Table 2. Size classification of clastic sedimentary particles and aggregates

Size, mm	Particle	Aggregate
Greater than 256	Boulder	Gravel conglomerate (psephite, rudite)
256–64	Cobble	Breccia (angular)
64–4	Pebble	
4–2	Coarse sand	Sandstone (psammite, arenite)
2–1/16	Sand	
1/16–1/256	Silt	Siltstone (pelite, lutite)
Less than 1/256	Clay	Clay Shale

Chemical sedimentary rocks are those precipitated from ocean, lake, and ground waters The most important ones are shown in the following list

Rock	*Chief mineral*
Chert	Chalcedony SiO_2
Limestone	Calcite $CaCO_3$
Travertine (spring deposit)	Calcite $CaCO_3$
Dolomite	Dolomite $CaMg(CO_3)_2$
Phosphorite	Apatite $Ca_{10}(PO_4)_6(CO_3)F_2$
Salines (evaporites)	
Rock salt	Halite NaCl
Rock anhydrite	Anhydrite $CaSO_4$
Rock gypsum	Gypsum $CaSO_4 \cdot 2H_2O$

Organic sedimentary rocks include (1) siliceous types made up of opaline tests of diatoms diatomite or radiolaria radiolarite (2) calcareous types, consisting of calcite shells—shell limestone and coquina and (3) carbonaceous types—coal and other accumulations of altered plant debris *See* SEDIMENTARY ROCKS SEDIMENTATION (GEOLOGY)

Metamorphic rocks Metamorphic rocks owe their complexity of composition and texture not only to the existence of several types of metamorphism but also to the application of these types under different intensities to a variety of parent

Table 1 Simplified classification of major igneous rocks on the basis of composition and texture

	SiO₂-rich (acidic) ←			→ SiO_2-poor (basic)	
	Light colored ←	Gray	Dark colored		→ Black
Mineral composition	Quartz potash feldspar biotite	Potash feldspar biotite or amphibole	Sodic plagioclase hornblende or augite	Augite olivine hypersthene calcic plagioclase	Olivine enstatite augite
Intrusive					
Medium grained	Granite*	Syenite	Diorite	Gabbro	Peridotite
Extrusive					
Fine-grained to aphanitic	Rhyolite ←—	Trachyte Felsite	Andesite —→	Basalt	
Porphyritic	Rhyolite porphyry	Trachyte porphyry	Andesite porphyry	Basalt porphyry	
Glassy	Obsidian				
Vesicular	Pumice			Scoria	
Fragmental	Tuff and agglomerate of each type				

* Exceptionally coarse-grained rock of general granitic composition is pegmatite

Table 3 Simplified classification of metamorphic rocks, with selected examples

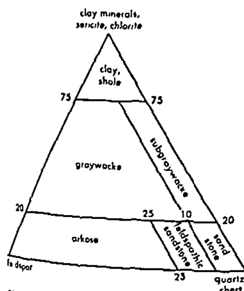
Parent rocks	Contact metamorphism	Regional metamorphism		
		Low grade	Intermediate grade	High grade
Sandstone arkose	Quartzite (quartz)	Quartzite and quartz feldspar gneiss		
Shale	Hornfels (andalusite, cordierite)	Slate, phyllite, (chlorite, muscovite)	Mica schist (biotite, garnet, kyanite, staurolite)	Sillimanite gneiss (sillimanite, almandite)
Limestone	Marble (calcite)	Marble and calc-silicate gneiss		
		(Calcite, tremolite)	(Calcite, wollastonite)	(Calcite, diopside, anorthite)
Basalt	Hornfels (plagioclase, hypersthene)	Greenschist (chlorite, albite, epidote)	Amphibole schist, (plagioclase, actinolite)	Amphibolite (andesine, hornblende, garnet)

rocks, thus, both sedimentary and igneous rocks [illegible]

Physical properties and behavior. When stresses are applied to rocks, either natural (those active in mountain building) or man made, resulting from loading with structures (such as dams) deformation (strain) may result Rocks subjected to stress normally undergo three deformation stages (1) elastic—the rock returns to its original size or shape upon withdrawal of stress, (2) plastic—beyond a limiting stress (elastic limit) there is only partial restoration upon stress removal, and (3) fracture—breakage with further increase in stress. With increases in confining pressure (load of overlying rocks) and temperature, the interval between the elastic limit and fracture increases Thus rocks that behave as brittle substances near the earth's surface failing by fracturing, will, at depth, be deformed plastically by solid flow The effects of stress depend on physical properties (Table 4) *See* ENGINEERING GEOLOGY, ROCK MECHANICS, STRUCTURAL GEOLOGY

Table 4 Physical properties of some common rocks

Rock	Specific gravity	Porosity %	Compressive strength psi	Tensile strength psi
Igneous				
Granite	2.67	1	30 000–50 000	500–1000
Basalt	2.75	1	25 000–30 000	
Sedimentary				
Sandstone	2.1–2.5	5–30	5 000–15 000	100–200
Shale	1.9–2.6	7–25	5 000–10 000	
Limestone	2.2–2.6	2–20	2 000–20 000	400–450
Metamorphic				
Marble	2.5–2.8	0.5–2	10 000–30 000	700–1000
Quartzite	2.5–2.6	1–2	15 000–40 000	
Slate	2.6–2.8	0.5–5	15 000–30 000	



Classification of psammitic and pelitic rocks based on proportions of three most common clastic minerals: quartz, feldspars, and clays—excluding calcilutites and calcarenites (From E W. Heinrich, Microscopic Petrography, McGraw Hill, 1956)

The rock cycle. Igneous rocks exposed at the earth's surface are subject to weathering, which alters them chemically and physically Such material when transported deposited and consolidated becomes sedimentary rock, which, through heat and *pressure, may be converted to metamorphic rock* Both sedimentary and metamorphic rocks also may be weathered and transformed into younger sediments Deeply buried metamorphic rocks may be remelted to yield new igneous material *See* LITHOSPHERE, GEOCHEMISTRY OF; *see also* PETROGRAPHY, PETROLOGY, STONE AND STONE PRODUCTS [L W H]

Rock (age determination)

The determination of the time of formation of crustal rocks Absolute rock age determination became possible in principle, with the discovery of radioactivity Prior to this discovery, the methods of stratigraphy and paleontology provided the only means of determining time of rock formation These methods were and still are useful in establishing the sequence of events and in defining the successions of strata They aid in piecing together the geologic time scale, but provide only a relative geochronology *See* DATING METHODS, GEOLOGICAL TIME SCALE, RADIOACTIVITY

Theory. All modern isotopic methods of age determination depend on the following principles

Organic (biogenic)
Metamorphic
Cataclastic
Contact metamorphic and pyrometasomatic
Regional metamorphic (dynamothermal)
Hybrid
Metasomatic
Migmatitic

Exceptions to the requirement that rocks consist of minerals are obsidian, a volcanic rock consisting of glass, and coal, a sedimentary rock which is a mixture of organic compounds *See* COAL, VOLCANIC GLASS

Igneous rocks Igneous rocks are those that have solidified from a molten condition The parent material is magma—a natural, hot, mutual solution of silicates with minor amounts of water and other volatiles Igneous rocks are divided into those which crystallized before magma reached the earth's surface (intrusive rocks) and those that solidified at the surface, some as layers of lava (the extrusive flow rocks) and others as pyroclastic debris in explosive eruption (Table 1) *See* MAGMA, PYROCLASTIC ROCKS

Chief elements in igneous rocks are oxygen (O), silicon (Si) aluminum (Al), iron (Fe), magnesium (Mg) calcium (Ca), sodium (Na), and potassium (K) Compositions range from about 40% SiO_2 (peridotites) to as much as 70% SiO_2 (granites) Silica poor rocks contain relatively large amounts of Ca, Mg and Fe^2 (basic rocks) whereas silica rich types contain larger amounts of Na and K (acidic rocks) *See* IGNEOUS ROCKS

Sedimentary rocks Clastic sedimentary rocks (consisting of mechanically transported particles) are subdivided on the basis of particle size (Table 2) Those having intermediate and fine-grain sizes are further subdivided on the basis of composition (Fig 1) Other significant clastic rocks are those consisting of detrital calcite the calcarenites and calcilutites *See* CALCARENITE

Textures of clastic rocks derive from grain size sorting form and arrangement Form includes sphericity (shape) the degree to which a particle approximates a sphere, and roundness the measurement of the sharpness of edges and corners Only glacial sedimentary rocks (tillite) do not show layering or stratification

Table 2. Size classification of clastic sedimentary particles and aggregates

Size, mm	Particle	Aggregate
Greater than 256	Boulder	Gravel conglomerate (psephite rudite)
256–64	Cobble	Breccia (angular)
64–4	Pebble	
4–2	Coarse sand	Sandstone (psammite, arenite)
2–1/16	Sand	
1/16–1/256	Silt	Siltstone (pelite lutite)
Less than 1/256	Clay	Clay Shale

Chemical sedimentary rocks are those precipitated from ocean, lake, and ground waters The most important ones are shown in the following list

Rock	*Chief mineral*
Chert	Chalcedony, SiO_2
Limestone	Calcite $CaCO_3$
Travertine (spring deposit)	Calcite $CaCO_3$
Dolomite	Dolomite $CaMg(CO_3)_2$
Phosphorite	Apatite, $Ca_{10}(PO_4)_6(CO_3)F_2$
Salines (evaporites)	
Rock salt	Halite, NaCl
Rock anhydrite	Anhydrite $CaSO_4$
Rock gypsum	Gypsum $CaSO_4 \cdot 2H_2O$

Organic sedimentary rocks include (1) siliceous types made up of opaline tests of diatoms diatomite, or radiolaria radiolarite (2) calcareous types consisting of calcite shells—shell limestone and coquina, and (3) carbonaceous types—coal and other accumulations of altered plant debris *See* SEDIMENTARY ROCKS, SEDIMENTATION (GEOLOGY)

Metamorphic rocks Metamorphic rocks owe their complexity of composition and texture not only to the existence of several types of metamorphism but also to the application of these types under different intensities to a variety of parent

Table 1 Simplified classification of major igneous rocks on the basis of composition and texture

	SiO_2-rich (acidic) ←				→ SiO_2-poor (basic)
	Light colored ←		Gray	Dark colored	→ Black
Mineral composition	Quartz potash feldspar biotite	Potash feldspar biotite or amphibole	Sodic plagioclase hornblende or augite	Augite olivine hypersthene calcic plagioclase	Olivine enstatite augite
Intrusive					
Medium grained	Granite*	Syenite	Diorite	Gabbro	Peridotite
Extrusive					
Fine-grained to aphanitic	Rhyolite ←	Trachyte — Felsite —	Andesite →	Basalt	
Porphyritic	Rhyolite porphyry	Trachyte porphyry	Andesite porphyry	Basalt porphyry	
Glassy	Obsidian				
Vesicular	Pumice			Scoria	
Fragmental	Tuff and agglomerate of each type				

* Exceptionally coarse-grained rock of general granitic composition is pegmatite

Table 2. Comparison of isotopic ages, 10^6 years, obtained by different methods*

Locality	K Ar	Rb-Sr	U^{238}-Pb^{206}	U^{235}-Pb^{207}
Beartooth Mts. Mont.	2,00 M	2 40 M	2600 U	2640 U
Keystone, S D	15 0 M	1690 M	1610 U	1620 U
Wilberforce Ont.	960 B	1020 B	1020 U	1020 U
Goodhouse South Africa	970 B		930 Mo	915 Mo
Wichita Mts. Okla.	460 B	510 B	517 Z	525 Z
Redstone, N H	168 B	190 B	18" Z	184 Z

* Decay constants same as in Table 1 B = biotite M = muscovite, Mo = monazite, U = uraninite, Z = zircon

samples and have permitted the construction of absolute geologic history in many areas of the world. These methods can span all of earth history, although the analytical errors in the K Ar method for ages below 10 years become rather large All other methods require further research before they can be established as useful geochronometers

The U Pb Rb-Sr, and K Ar methods are discussed in some detail later Table 2 gives the isotopic ages obtained by the three primary chronometers on unaltered rocks of various ages. The results agree within the uncertainties of analysis and decay constants For the C^{14} method *see* RADIOCARBON DATING.

Rubidium strontium method The radioactive transformation of Rb^{87} to Sr^{87} by β decay provides one of the most reliable and useful isotopic geochronometers, particularly for older rocks The

[illegible]

negligible radiation damage, so that mineral phases which can be dated by this method are intact even in the oldest rocks. The Rb-Sr method has been applied successfully to muscovite and biotite in a great variety of igneous and metamorphic rocks and to all potash micas pollucite rhodizite amazonite, and perthite in pegmatites There are many other potassium minerals in *igneous and metamorphic* rocks that may be used including orthoclase, sanidine leucite and phlogopite provided the ratio *of* rubidium to common strontium is large enough for the age involved Glauconite appears promising for dating sedimentary rocks. The ratio of rubidium *to* common strontium in some stony meteorites is high enough for age determinations to be made on the whole rock

Assuming a half life of $5\ 0 \times 10^{10}$ years which is probably accurate to at least $\pm 6\%$ the age formula for the Rb Sr method is

$$T = 7.2 \times 10^4 \ln \left[\frac{Sr^{87*}}{Rb^{87}} + 1\right]$$

where T is the isotopic age in millions of years, Rb^{87} and Sr^{87*} are the number of radioactive rubidium and radiogenic strontium atoms respectively The isotopic age is the true age if no alteration has occurred during the history of the mineral The total rubidium and strontium contents of the sample are obtained with accuracies of 1–3% by routine, but sophisticated, isotope dilution techniques. The isotopic composition of the strontium in the sample must be analyzed separately The Rb^{87} can then be readily obtained from the known isotopic composition of natural rubidium, that is, $Rb^{85} = 72.15\%$ and $Rb^{87} = 27\ 85\%$ The calculation of the radiogenic Sr^{87*} content is more complex because all rubidium bearing minerals also contain measurable common strontium which was incorporated at the time of mineral formation The isotopic composition of strontium in modern ocean water is $Sr^{88} = 82.5\%$, $Sr^{87} = 7\ 02\%$ and $Sr^{86} = 9\ 85\%$ In the rock forming environment the isotopic composition of the common strontium has generally not changed significantly with time, so that the Sr^{86} abundance in the sample whose age is to be determined can be used to determine the fraction of the total Sr^{87} that was incorporated at the time of mineral formation The difference is the radiogenic Sr^{87*} In the event that there is a question about the isotopic composition of the common strontium incorporated into the rubidium rich mineral at the time of formation a rubidium poor but strontium rich phase such as plagioclase or apatite may be analyzed isotopically for strontium

It is evident that as the ratio of rubidium to common strontium in a rock increases the error in the determination of the radiogenic Sr^{87*} will increase Thus in order to obtain an isotopic age that is analytically accurate to $\pm 5\%$ assuming that the isotopic composition of the strontium can be measured to $\pm 0.5\%$ the ratio of rubidium to common strontium would have to be at least 10 for a mineral 1 800 000 000 years old but at least 100 for a mineral 180 000 000 years old

The Sr^{87*}/Rb^{87} ratio appears to remain unaffected in feldspar unless recrystallization occurs In biotite however there is some evidence that alteration may take place at lower temperatures possibly involving base exchange phenomena If *Rb-Sr ages on cogenetic mica and feldspar agree* it is strong evidence of a real date of the last metamorphic or igneous event

Potassium argon method The radioactive isotope of potassium K^{40} decays by β emission to Ca^{40} and by K electron capture to Ar^{40} The decay to Ca^{40} has only restricted value in geochronology because common calcium is largely Ca^{40} and most potassium minerals contain significant amounts of calcium The argon part of the decay is particularly attractive because potassium minerals appear to form without incorporating any primary argon from their environment and the analytical methods for argon determination are extremely sensitive The widespread occurrence of potassium minerals suggests that the method can be applied to virtually all rock complexes Using the decay constants for the separate branches of the disintegration of K^{40}, $\lambda_e = 0\ 584 \times 10^{-4}/10^6$ years and $\lambda_\beta = 4\ 72 \times 10^{-4}/10^6$ years the age equation becomes

$$T \text{ (in } 10^6 \text{ years)} = 1885 \ln (1 + 9\ 10\ Ar^{40}/K^{40})$$

The average rate of disintegration of a large number of radioactive atoms of a given kind is constant. The mathematical relationship is $-dN/dt = \lambda N$, where N is the number of atoms of this kind present and λ is the disintegration constant. This rate can be accurately measured in the laboratory by determining the number of atoms decaying per unit time for a given total number of atoms. The rate is not altered by any physical or chemical conditions to which a crustal rock could have been subjected since its origin. In the simplest case a mineral containing radioactive parent atoms A but not stable daughter atoms B, crystallizes from a silicate melt during the formation of an igneous rock. The longer the time interval between mineral formation and measurement, the larger the ratio of B/A. The age relationship is given precisely by

$$T = \frac{1}{\lambda} \ln \frac{B+1}{t}$$

where T is the age (generally given in millions of years). The rate of radioactive disintegration is also given by the half life of the isotope, that is, the time required for half of the existing number of atoms to disintegrate. The half life ($t_{1/2}$) is related to the decay constant by $t_{1/2} = .693/\lambda$. For practical purposes the half life of a radioactive isotope must be the same order of magnitude as the time span to be measured. Thus isotopes of interest for geochronology must have half lives ranging from thousands of years (for the study of recent geological and archeological events) to hundreds of millions of years (for ancient rocks and the age of the earth). Providing the parent (A) and daughter (B) atoms are quantitatively separated at the time of mineral formation and that the mineral has remained a closed chemical system during its history, an absolute age can be obtained the accuracy of which is only limited by the uncertainties in the chemical and isotopic analyses of A and B and in the half life. Under the most favorable conditions the age determinations can be made to within a few per cent even for a rock thousands of millions of years old.

The early phase of quantitative age determination based on radioactivity occurred between 1900 and 1938. During this period work was restricted to the measurement of lead uranium ratios in uranium minerals and in helium uranium ratios in a variety of mineral and rock types. This pioneer work established the order of magnitude of the geologic time scale and provided the first age measurements for many geologic provinces. It suffered from relatively crude analytical methods that were available, inadequate knowledge of the nuclear phenomena involved, and the absence of reliable criteria for identifying chemical alteration or the incorporation of daughter product of the isotopic clocks at the time of mineral formation.

Modern isotopic geochronometry began with the first comprehensive and precise measurement of the lead isotopes in uranium and lead minerals by A. O. Nier and his coworkers in 1939. The development of the instrumentation and analytical procedures for microassay by the isotope dilution method and the precise measurement of the isotopic composition made possible the discovery and application of a number of isotopic chronometers apart from the uranium lead system. The possibility of obtaining independent age estimates from different mineral phases provided the necessary criteria for demonstrating a closed chemical system and for detecting primary contaminants. In addition, valuable geochemical information could be obtained from partially open mineral systems.

The most important age determination methods are summarized in Table 1. The U-Pb (uranium lead), Rb-Sr (rubidium strontium), K-Ar (potassium argon), and C^{14} (carbon 14) methods have been shown to yield reliable results on suitable

Table 1 Methods of quantitative geochronometry

Method	$T_{1/2}$ years	Effective range in years	Applicable minerals or rocks
Rb^{87}-Sr^{87}	5.0×10^{10}	T_e*–10^8	Muscovite, biotite, K feldspar, lepidolite, glauconite
K^{40}-Ar^{40}	1.3×10^{9}†	T_e–10^4	Muscovite, biotite, glauconite
U^{238}-Pb^{206}	4.5×10^{9}	T_e–10^7	Uraninite, monazite, zircon, black shale
U^{235}-Pb^{207}	0.71×10^{9}		
Th^{232}-Pb^{208}	1.39×10^{10}		
C^{14}	5.6×10^{3}	50 000–present	Carbon bearing materials once in contact with the atmosphere-biosphere system
U-Th-He		T_e–10^6(?)	Possibly sulfides and magnetite
K^{40}-Ca^{40}	1.3×10^{9}†	T_e–10^8(?)	Sylvite
Pb^{207}-Pb^{206}		T_e–10^8	Galena, lead in pyrite
Radiation damage		Uncertain	Zircon, samarskite
Cl^{36}	3×10^{5}	Uncertain	Chlorine-rich rocks exposed to cosmic ray neutrons
H^{3}	12	50–present	Ground water
Ionium	8×10^{4}	4×10^{5}–present	Deep-sea sediments
Re^{187}-Os^{187}	6×10^{10}(?)	T_e–?	Old molybdenites
Lu^{176}-Hf^{176}	2×10^{10}	T_e–?	Old rare-earth minerals

* T_e is the age of the earth (about 4.6×10^9 years). † Total half life.

Table 2. Comparison of isotopic ages, 10^6 years, obtained by different methods*

Locality	K-Ar	Rb-Sr	U^{238}-Pb^{206}	U^{235}-Pb^{207}
Beartooth Mts., Mont.	2500 M	2[illegible]40 M	2600 U	2640 U
Keystone S D	13[illegible]0 M	1690 M	1610 U	1620 U
Wilberforce, Ont.	960 B	1020 B	1020 U	1020 U
Goodhouse South Africa	9[illegible]0 B		930 Mo	915 Mo
Wichita Mts. Okla	460 B	510 B	517 Z	525 Z
Redstone, N H	168 B	190 B	187 Z	184 Z

*Decay constants same as in Table 1 B = biotite M = muscovite, Mo = monazite, U = uraninite, Z = zircon.

samples and have permitted the construction of [illegible] for ages below 10 years become rather large All other methods require further research before they can be established as useful geochronometers

The U-Pb Rb-Sr, and K Ar methods are discussed in some detail later Table 2 gives the isotopic ages obtained by the three primary chronometers on unaltered rocks of various ages The results agree within the uncertainties of analysis and decay constants. For the C^{14} method *see* RADIOCARBON DATING

Rubidium strontium method The radioactive transformation of Rb^{87} to Sr^{87} by β-decay provides one of the most reliable and useful isotopic geochronometers particularly for older rocks The method can be applied to Cenozoic rocks providing a sufficiently high Rb/Sr ratio is present as in lepidolite. Ordinary biotite in granites as young as Mesozoic can be dated by the Rb-Sr transformation The soft energy of the Rb^{87} β particles produces negligible radiation damage so that mineral phases which can be dated by this method are intact even in the oldest rocks. The Rb-Sr method has been applied successfully to muscovite and biotite in a great variety of igneous and metamorphic rocks and to all potash micas, pollucite rhodizite, amazonite, and perthite in pegmatites There are many other potassium minerals in igneous and metamorphic rocks that may be used including orthoclase, sanidine leucite and phlogopite provided the ratio of rubidium to common strontium is large enough for the age involved Glauconite appears promising for dating sedimentary rocks The ratio of rubidium to common strontium in some stony meteorites is high enough for age determinations to be made on the whole rock

Assuming a half life of 5 0 × 10^{10} years which is probably accurate to at least ±6%, the age formula for the Rb-Sr method is

$$T = 7\,2 \times 10^4 \ln \left[\frac{Sr^{87*}}{Rb^{87}} + 1\right]$$

where T is the isotopic age in millions of years Rb^{87} and Sr^{87*} are the number of radioactive rubidium and radiogenic strontium atoms respectively The isotopic age is the true age if no alteration has occurred during the history of the mineral The total [illegible] strontium contents of the sample are obtained with accuracies of 1-3% by routine, but sophisticated, isotope dilution techniques The isotopic composition of the strontium in the sample must be analyzed separately The Rb^{87} can then be readily obtained from the known isotopic composition of natural rubidium, that is, Rb^{85} = 72.15% and Rb^{87} = 27 85% The calculation of the radiogenic Sr^{87*} content is more complex because all rubidium bearing minerals also contain measurable common strontium which was incorporated at the time of mineral formation The isotopic composition of strontium in modern ocean water is Sr^{88} = 82.5%, Sr^{87} = 7 02%, and Sr^{86} = 9 85% In the rock forming environment the isotopic composition of the common strontium has generally not changed significantly with time, so that the Sr^{86} abundance in the sample whose age is to be determined can be used to determine the fraction of the total Sr^{87} that was incorporated at the time of mineral formation The difference is the radiogenic Sr^{87*} In the event that there is a question about the isotopic composition of the common strontium incorporated into the rubidium rich mineral at the time of formation, a rubidium poor but strontium rich phase such as plagioclase or apatite may be analyzed isotopically for strontium

It is evident that as the ratio of rubidium to common strontium in a rock increases, the error in the determination of the radiogenic Sr^{87*} will increase Thus in order to obtain an isotopic age that is analytically accurate to ±5% assuming that the isotopic composition of the strontium can be measured to ±0.5% the ratio of rubidium to common strontium would have to be at least 10 for a mineral 1 800 000 000 years old but at least 100 for a mineral 180 000 000 years old

The Sr^{87*}/Rb^{87} ratio appears to remain unaffected in feldspar unless recrystallization occurs. In biotite however, there is some evidence that alteration may take place at lower temperatures possibly involving base exchange phenomena If Rb-Sr ages on cogenetic mica and feldspar agree, it is strong evidence of a real date of the last metamorphic or igneous event

Potassium argon method The radioactive isotope of potassium K^{40} decays by β emission to Ca^{40} and by *K*-electron capture to Ar^{40} The decay to Ca^{40} has only restricted value in geochronology because common calcium is largely Ca^{40} and most potassium minerals contain significant amounts of calcium The argon part of the decay is particularly attractive because potassium minerals appear to form without incorporating any primary argon from their environment and the analytical methods *for argon determination are extremely sensitive* The widespread occurrence of potassium minerals suggests that the method can be applied to virtually all rock complexes Using the decay constants for the separate branches of the disintegration of K^{40} λ_e = 0 584 × $10^{-4}/10^6$ years and λ_β = 4 72 × $10^{-4}/10^6$ years the age equation becomes

$$T \text{ (in } 10^6 \text{ years)} = 1885 \ln (1 + 9\,10\, Ar^{40}/K^{40})$$

The K^{40} isotopic abundance is constant (0.0119%) in natural potassium so that this isotope may be determined by standard wet-chemical methods. The Ar^{40} is determined by the isotope dilution method after it is released quantitatively from the mineral by fusion. Correction is made for contamination by Ar^{40} from air by monitoring the Ar^{36} in the sample gas. Isotopic ages may be determined by this method over most of geologic time with an accuracy of a few per cent. For young minerals the atmospheric correction becomes a limiting factor.

At low temperatures mica appears to retain essentially all of its radiogenic argon (Table 2). At elevated temperatures diffusion may cause partial loss of radiogenic argon. This has been observed in field tests along the border of a younger metamorphic belt that is superimposed on a preexisting basement.

Other minerals such as feldspar, lepidolite, and glauconite do not appear to hold all of the radiogenic argon. It may be, however, that criteria can be developed for estimating the degree of retention for various structural types. Measurements on these minerals and whole rocks (particularly basalts and siliceous effusives) yield minimum ages only, but these may be of great value for certain geological problems.

Uranium-lead method. The method based on the decay of the thorium and uranium to lead is the oldest and most elegant; it involves two or three radioactive isotopes, U^{238}, U^{235} and Th^{232}, to three distinct isotopes of lead. In pitchblende the thorium content is negligible but in most other uranium or thorium minerals both elements are present in sufficient quantities for analysis. Each of the above isotopes decays through a series of 8–12 isotopes until a stable lead isotope is produced, Pb^{206}, Pb^{207}, and Pb^{208}, respectively. The intermediate isotopes all have much shorter half lives than the parent isotopes, so that except for *alteration effects* the chronometers may be considered as simple parent-daughter decay systems of U^{238} to Pb^{206}.

The presence of two uranium isotopes with different half lives and different chemical intermediates with distinctive nuclear properties provides a mutual check on the reliability of the ages obtained. Any chemical alteration in the system will affect the apparent uranium-to-lead isotopic ratios in such a way that the calculated ages will differ. Such a discordance can be used to evaluate the nature, time, and extent of alteration if sufficient data are available. Concordance among these ratios is strong evidence that a true age has been obtained.

The age determination consists of analyzing the sample for total uranium, thorium, and lead by the isotope dilution method and for its lead isotopic composition. There is only one significant isotope of thorium (Th^{232}) and two of uranium (U^{238} and U^{235}) in constant proportion so that these quantities are readily calculated from the chemical analysis. The lead presents a more complex problem because nearly all uranium and thorium minerals incorporate at least small quantities of common rock lead at the time of formation. This rock lead contains Pb^{204}, Pb^{206}, Pb^{207}, and Pb^{208}. All of these isotopes were present in the primeval lead of the earth but additional Pb^{206}, Pb^{207}, and Pb^{208} have been added throughout earth history as a result of the radioactive decay of U^{238}, U^{235} and Th^{232} in the crust and mantle. The ratios of Pb^{206}/Pb^{204}, Pb^{207}/Pb^{204}, and Pb^{208}/Pb^{204} are therefore a function of time. They have increased only slowly and therefore it is generally sufficient to use the average crustal value at the approximate time of mineral formation in order to make the correction for the incorporated lead. If necessary, a more precise correction can be made by analyzing the lead isotopic composition in a uranium-free mineral that is cogenetic with the uranium mineral. Thus if the lead isotopic ratios in the sample and in the contaminating common lead are known, the quantity of the radiogenic isotopes Pb^{206}, Pb^{207}, and Pb^{208} may be derived. The ages are then calculated using equations analogous to that shown above for the simple decay of Rb^{87} to Sr^{87}. See ISOTOPE DILUTION TECHNIQUES; LEAD ISOTOPES, GEOCHEMISTRY OF.

In some cases, such as those in Table 2, the isotopic ages obtained from the uranium isotopes agree within the experimental error. This appears generally true for large fresh uraninite crystals from pegmatites in areas that have not been subjected to any later thermal effects. In many cases, however, significant discordance occurs. Most often this is due to selective lead loss, but loss of intermediate decay products may also play a significant role. By examining different minerals and different samples of the same mineral it is possible to reconstruct the geochemical history as well as the original age of the mineral.

Although the original use of the U-Pb method was to date pegmatite crystals or pitchblende veins in metallic ores, the most important application appears to be to accessory zircon crystals in igneous and metamorphic rocks. Some attempts have been made to apply the method to date uranium-rich black shales but the high common lead content and ease of alteration make a precise determination difficult.

The variation of the average lead isotopic composition in the crust has been useful in obtaining the approximate ages of ore deposits. The U-Pb method has also been found applicable to the age of meteorites. See EARTH (AGE OF); METEORITE; see also RADIOACTIVE MINERALS. [J.L.K.]

Bibliography: L. T. Aldrich and G. W. Wetherill, Geochronology by radioactive decay, *Ann. Rev. Nuclear Sci.*, 8:257–298, 1958; J. L. Kulp, *Quantitative Geochronometry*, 1962.

Rock magnetism

The induced and permanent magnetization of certain rocks, from which inferences concerning the geomagnetic field in past geological times and

the relative movements of the poles and the conti nents have been made

The magnetization of rocks results from the iron oxide minerals present, to a few per cent in many rocks. Two groups are important the magnetite (Fe_3O_4) ulvospinel (Fe_2TiO_4) solid solution series, and the hematite (Fe_2O_3) ilmenite ($FeTiO_3$)

tion of air or ground geomagnetic surveys

Primary magnetization Three types of remanent magnetization may be acquired during the process of formation of an igneous or sedimentary rock

Thermoremanent magnetization The permanent or natural remanent magnetization of igneous rocks, such as lavas, sills, and dykes, is due to the magnetization of these minerals during cooling in the geomagnetic field from the temperatures at which they solidify, which are usually above their Curie points.

Depositional and chemical magnetization In sediments the magnetization may arise through the orientation of the detrital iron oxide grains by the geomagnetic field during deposition Alternatively it may arise during chemical change, as in the growth of grains of hematite in the red sandstones, during or perhaps slightly later than deposition The former type is known as depositional the latter as chemical magnetization

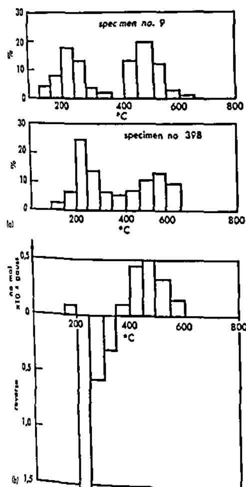


Fig. 1 Graphs showing partial thermoremanent magnetization (a) Icelandic specimens of Tertiary plateau basalt with reverse natural permanent magnetization For easier comparison the intensities are expressed in per cents of the sum for no 9 it is 13.2×10^{-3} gauss and for no 398 5.6×10^{-3} gauss The partial thermoremanent magnetization is in the direction of the field for all temperature intervals (b) A specimen of dacite pumice from Mt Haruna Japan Normal denotes that the magnetization is in the direction of the external field "reverse" that it is in the opposite direction (After J Hospers in S K Runcorn Magnetization of Rocks in S Fluegge, ed, Handbuch der Physik, vol 47 1956)

Secondary magnetization. On these original magnetizations is often superposed a secondary magnetization acquired in more recent times This is due in part to magnetically unstable iron oxide grains which may be in the original rock or may result from weathering According to the theory of L Neel these grains are either too small or too finely divided by intergrowths of different chemical composition to retain a permanent magnetization indefinitely and if placed in zero magnetic field will gradually lose their magnetization logarithmically with time Alternatively such grains will according to the same law acquire a magnetization parallel to the field in which they lie This is known as viscous magnetization Rocks are found with varying amounts of this secondary magnetization, sometimes changing appreciably during laboratory storage but more often possessing a relatively stable component along the present geomagnetic field

Secondary magnetization of this type may sometimes be proved to be absent This is best done by showing that specimens of the rock tilted folded or dispersed in a conglomerate since the time of the original magnetization have consistent directions of magnetization when allowance is made for these geological processes Secondary magnetization can sometimes be removed by the process of ac demagnetization or by heating to a few hundred degrees and cooling in zero magnetic field Measurement of the direction of the original component of magnetization making allowance if necessary for geological tilting enables changes in the geomagnetic field to be traced throughout the geological record in different places

Magnetization of metamorphic rocks Little is yet known of this phenomenon but components of magnetization may be acquired through different processes at different stages in the history of the rock

Geomagnetic secular variation In historic times such variation has been studied with archaeological specimens, dated varve clays and lavas from dated eruptions (*see* GEOMAGNETISM) The first two investigations show that the geomagnetic field has had a secular variation similar to that observed in the last few hundred years The third method ap

The K^{40} isotopic abundance is constant (0 0119%) in natural potassium so that this isotope may be determined by standard wet chemical methods The Ar^{40} is determined by the isotope dilution method after it is *released quantitatively* from the mineral by fusion Correction is made for contamination by Ar^{40} from air by monitoring the Ar^{36} in the sample gas Isotopic ages may be determined by this method over most of geologic time with an accuracy of a few per cent For young minerals the atmospheric correction becomes a limiting factor

At low temperatures mica appears to retain essentially all of its radiogenic argon (Table 2) At elevated temperatures diffusion may cause partial loss of radiogenic argon This has been observed in field tests along the border of a younger metamorphic belt that is superimposed on a preexisting basement

Other minerals such as feldspar lepidolite and glauconite do not appear to hold all of the radiogenic argon It may be, however that criteria can be developed for estimating the degree of retention for various structural types Measurements on these minerals and whole rocks (particularly basalts and siliceous effusives) yield minimum ages only but these may be of great value for certain geological problems

Uranium lead method The method based on the decay of the thorium and uranium to lead is the oldest and most elegant it involves *two or* three radioactive isotopes U^{238} U^{235} and Th^{232} to three distinct isotopes of lead In pitchblende the thorium content is negligible but in most other uranium or thorium minerals both elements are present in sufficient quantities for analysis Each of the above isotopes decays through a series of *8–12 isotopes until a stable lead isotope is pro*duced Pb^{206} Pb^{207} and Pb^{208} respectively The intermediate isotopes all have much shorter half lives than the parent isotopes so that except for *alteration effects the* chronometers may be considered as simple parent daughter decay systems of U^{238} to Pb^{206}

The presence of two uranium isotopes with different half lives and different chemical intermediates with distinctive nuclear properties provides a mutual check on the reliability of the ages obtained Any chemical alteration in the system will affect the apparent uranium to-lead isotopic ratios in such a way that the calculated ages will differ Such a discordance can be used to evaluate the nature time and extent of alteration if sufficient data are available Concordance among these ratios is strong evidence that a true age has been obtained

The age determination consists of analyzing the sample for total uranium thorium and lead by the isotope dilution method and for its lead isotopic composition There is only one significant isotope of thorium (Th^{232}) and two of uranium (U^{238} and U^{235}) in constant proportion so that these quantities are readily calculated from the chemical analysis The lead presents a more complex problem because nearly all uranium and thorium minerals incorporate at least small quantities of common rock lead at the time of formation This rock lead contains Pb^{204} Pb^{206} Pb^{207} and Pb^{208} All of these isotopes were present in the primeval lead of the earth but additional Pb^{206} Pb^{207} and Pb^{208} have been added throughout earth history as *a result of the* radioactive decay of U^{238} U^{235} and Th^{232} in the crust and mantle. The ratios of Pb^{206}/Pb^{204} Pb^{207}/Pb^{204} and Pb^{208}/Pb^{204} are there*fore a function* of time. They have increased only slowly and therefore it is generally sufficient to use the average crustal value at the approximate time *of mineral* formation in order to make the correction for the incorporated lead If necessary a more precise correction can be made by analyzing the *lead isotopic* composition in a uranium free mineral that is cogenetic with the uranium mineral Thus if the lead isotopic ratios in the sample and in the contaminating common lead are known the *quantity of* the radiogenic isotopes Pb^{206} Pb^{207} and Pb^{208} may be derived The ages are then calculated using equations analogous to that shown *above for* the simple decay of Rb^{87} to Sr^{87} *See* ISOTOPE DILUTION TECHNIQUES LEAD ISOTOPES, GEOCHEMISTRY OF

In some cases such as those in Table 2, the isotopic ages obtained from the uranium isotopes agree within the experimental error This appears generally true for large fresh uraninite crystals *from* pegmatites in areas that have not been subjected to any later thermal effects. In many cases, however significant discordance occurs. Most often this *is due to selective lead loss* but loss of intermediate decay products may also play a significant role By examining different minerals and different samples of the same mineral it is possible to *reconstruct the geochemical history as well as the* original age of the mineral

Although the original use of the U Pb method was to date pegmatite crystals or pitchblende veins in metallic ores the most important application appears to be to accessory zircon crystals in igneous and metamorphic rocks Some attempts have been made to apply the method to date uranium rich black shales but the high common lead content and ease of alteration make a precise determination difficult

The variation of the average lead isotopic composition in the crust has been useful in obtaining the approximate ages of ore deposits The U Pb method has also been found applicable to the age of meteorites *See* EARTH (AGE OF) METEORITE *see also* RADIOACTIVE MINERALS [J L K]

Bibliography L T Aldrich and G W Wetherill Geochronology by radioactive decay *Ann Rev Nuclear Sci* 8 257 298 1958 J L Kulp *Quantitative Geochronometry* 1962

Rock magnetism

The induced and permanent magnetization of certain rocks from which inferences concerning the geomagnetic field in past geological times and

nental mass but different from those of other periods and other continents. A plot of the path of t . . l log cal time can be drawn for

occurring in relatively late geological times ... basis of the results for Cenozoic times and from the theory of the geomagnetic field the mean field is assumed to have an axis coincident with the axis of rotation.

Polar wandering is therefore also inferred from the slow change of the direction of magnetization through the geological column (see POLAR WANDER ING)

Another representation of the data is useful The ba c a sumptions reduce to the following relation between the angle of inclination *I* to the horizontal of the mean paleomagnetic field and the ancient latitude λ of the site where the rock specimens were collected

$$\tan I = 2 \tan \lambda$$

... lec mag ... ridian ... a lati tude grid fixed to the axis of rotation may be inferred. Its motion in longitude is not determined

Paleomagnetic surveys of the continents indicate that North America and Europe, South America and Africa have drifted apart since early Mesozoic times and that India was south of the Equator in Mesozoic times A Wegener's reconstruction of the continents and the formation of ocean basins has thus received some support. *See* SUBMARINE TOPOGRAPHY

Table of geological age and formation of rocks yielding evidence of positions of Australia

Position (Fig 3)	Geological age	Formation
PPR	Pliocene Pleistocene, and Recent	Newer volcanics of Victoria
E	Lower Tertiary probably Eocene	Older volcanics of Victoria
J	Mesozoic, probably Jurassic	Dolerite sills of Tasmania
Tr	Triassic, probably Lower Triassic	Brisbane tuff
P_1	Permian, Upper Marine Series	Volcanics of Illawarra coast
P_1	Permian, Lower Marine Series	Volcanics of Hunter Valley
C	Upper Carboniferous	Kuttung red varvoid sediments
C		
D	Upper Carboniferous	Kuttung lavas
	Devonian, probably Lower Devonian	Ainslie volcanics
S	Upper Silurian	Mugga porphyry
ϵ_1	Middle Cambrian	Elder Mountain sandstone
ϵ_1	Lower Cambrian	Antrim Plateau basalts
Pre_{ϵ_3}	Top of Upper Proterozoic	Buldiva quartzite
Pre_{ϵ_2}	Upper Proterozoic	Mallagine lavas
Pre_{ϵ_1}	Lower part of Upper Proterozoic	Edith River volcanics

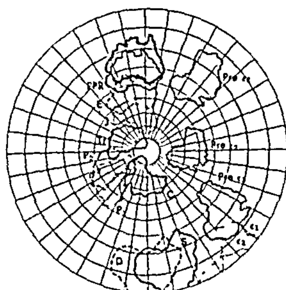


Fig 3 Movement of Australia relative to South Pole of rotation Heavy outline denotes present positions; see table for geological age and formation of rocks yielding evidence

The paleomagnetic observations could be interpreted without the aid of these far reaching hypotheses concerning the evolution of the earth's crust if it were assumed either that the original magnetization of rocks was controlled to an important degree by other causes such as flow in igneous rocks or the orienting action of water currents in sediments, or that the mean geomagnetic field in earlier geological times was not dipolar The field evidence is predominantly though perhaps not decisively against these alternative explanations

Many rocks have an anisotropic susceptibility, show magnetostriction and have very high coercivities Igneous rocks usually have intensities of magnetization between 10^{-3} and 10^{-4} gauss and sedimentary rocks between 10^{-6} and 10^{-7} gauss The intensity of the geomagnetic field at the time of magnetization is only one of many factors influencing the remanent magnetization yet it appears that the geomagnetic dipole moment could not have changed by orders of magnitude during geological time [S K R]

Bibliography P M S Blackett *Lectures on Rock Magnetism* 1956 D W Collinson et al, Palaeomagnetic investigations in Great Britain *Phil Trans Roy Soc London* vol 250 1957 T Nagata *Rock Magnetism* 1953 L Neel Some theoretical aspects of rock magnetism *Advances in Physics* 4(14) 191–243 1955 G D Nicholls The mineralogy of rock magnetism *Advances in Physics* 4(14) 113–190 1955 S K Runcorn, Rock magnetism geophysical aspects *Advances in Physics* 4(14) 245–291 1955

Rock mechanics

Rocks deformed in nature exhibit characteristics of flow recrystallization and the development of slaty cleavage These characteristics of rocks are

plied to the Mount Etna lavas has shown that the directions of magnetization of the lava flows agree with the observations of the field of the same date

Geomagnetic reversals. A characterizing aspect for most of Cenozoic times (the Tertiary and Quaternary periods), world wide paleomagnetic observations show the mean geomagnetic field (in the sampling of a series of rocks the averaging is done usually over 10^3–10^5 years) to be that of a geocentric dipole along the present geographical axis. However, successive parts of the stratigraphical *column have magnetizations of opposed directions*. These reversed magnetizations may result from anomalous properties of certain of the iron oxide minerals, which are ferrimagnetic, or through chemical changes subsequent to original magnetization, but probably they are mostly to be explained by the geomagnetic dipole reversing its polarity at rather irregular intervals of the order of 10^6 years

Polar wandering and continental drift. For the most part, results of paleomagnetic observations and speculations concerning them have led to interesting though still controversial interpretations. The paleomagnetic results from early Cenozoic, Mesozoic, Paleozoic, and late Precambrian *rocks show magnetizations which, if interpreted as* arising from a dipole field not along the present geographical axis, give nearly similar pole positions for any one geological period within a conti

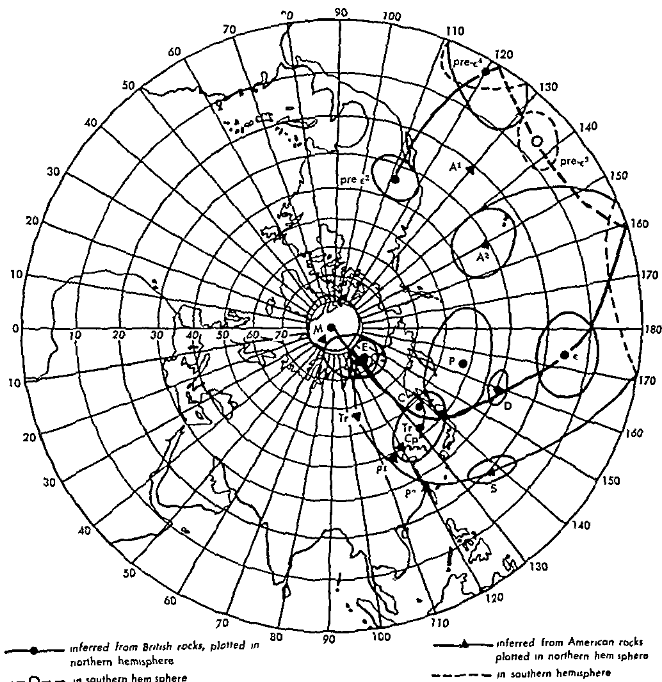


Fig 2 Pole positions and approximate path of North Pole (After K M Creer, E Irving, and S K Runcorn, Geophysical interpretation of palaeomagnetic directions from Great Britain, *Phil Trans Roy Soc London*, vol 250, 1957–1958)

strain rate has been ascertained

The study of naturally deformed rocks by the refined methods of structural petrology has resulted in a vast body of observational data In a few instances notably the study of H W Fairbairn and H. E. Hawkes, it has been possible to infer correctly the mechanism of deformation of crystals from a study of naturally deformed rocks alone Once the mechanism is known and confirmed in the laboratory then the data of structural petrology permit the reconstruction of tectonic events *See* STRUCTURAL PETROLOGY

Quartz is the one common mineral that has not yielded to laboratory or field studies Recent work by W F Brace with diamond indentations of

abo t an a axis in naturally deformed quartz, con

tallization in quartz, despite the fact that quartz is one of the most commonly deformed and recrystallized minerals in nature.

Analogy with metals When subjected to low stresses for long durations rocks deform at a very low rate analogous to creep in metals The mechanisms of flow in rock creep have not yet been determined but are presumed to be similar to the

flow in rocks deformed naturally, this mechanism is the same as that found in the laboratory

A specimen of Solenhofen limestone subjected to a compressive stress of 1 4 kb for 11 years at 25°C shortened at a logarithmically decreasing rate for the first three years then at an even slower rate. The compressive strain in the last eight years was only 0 002% The minimum equivalent viscosity for this interval was 10 ² poise Ordinary steel creeps at a very much faster rate than this

Rocks exhibit many other characteristics found in metals notably strain hardening and elastic after working Rocks and single crystals of the component grains also follow all the laws of behavior derived by the students of metals For example in a monomineralic rock deformed under conditions such that the mechanism of flow is intracrystalline gl de the law of maximum resolved shear stress is obeyed determining which of the several equivalent gl de systems in an individual crystal will function In the case of recrystallization it has been found that giant grain growth follows from achieving critical strain and not from very slow growth rates as presumed by some students of met amorphi m

The theory of deformation and recrystallization of crystals has not yet caught up with experiment, so that it is not possible to predict what will happen under a set of conditions that have not been explored experimentally The great gains in the future of rock mechanics must come *from a devel*

lel those with metals indicates that when theory is developed which will explain the behavior of metal crystals and aggregates it will also explain *the behavior of rocks.*

An understanding of rock mechanics is essential to elucidate the processes which mold the face of the earth and produce continents mountain ranges, earthquakes, and the structures which govern the formation of ore deposits and oil fields. *See* OROGENY, STRUCTURAL GEOLOGY, TECTONOPHYSICS, WARPING EARTH CRUST [D T G]

Bibliography S W Bailey R A Bell and C J Peng Plastic deformation of quartz in nature, *Bull Geol Soc Am* 69 1443 1466 1958, W F Brace Plastic deformation of quartz during indentation *Bull Geol Soc Am* 69 1539 1958, M J Buerger and F Washken Metamorphism of minerals *Am Mineralogist* 32 296–308 1947 H W Fairbairn and H E Hawkes Jr Dolomite orientation in deformed rocks *Am J Sci*, 239 617–632, 1941, *Rock Deformation* Geol Soc Am Memoir, 1960 F J Turner et al Deformation of yule marble pt 7 Development of oriented fabrics at 300°C–500°C *Bull Geol Soc Am*, 67 1259–1293, 1956

Rock salt

The mineral halite NaCl or common salt It crystallizes in the isometric system commonly in cubes It has perfect cubic cleavage a hardness of 2½ on Mohs scale and a specific gravity of 2 16 Rock salt is colorless or white but when impure may be various shades of yellow red blue or purple

Rock salt has both a broad geographic and geologic distribution It occurs in sedimentary rocks of all geologic ages since the Cambrian and is widely distributed throughout the world Salt commonly associated with gypsum anhydrite clay sylvite, carnallite and a variety of other soluble salts occurs in beds ranging from a few feet to over 1000 ft in thickness Deformation of deeply buried stratified deposits may result in local pluglike extrusions through the overlying strata forming salt domes The great storehouse of sodium chloride is in the oceans and it is from this source that the bedded deposits have been derived by evaporation in shallow basins Salt also occurs in salt seas and lakes, where its concentration is greater than in sea water *See* EVAPORITE (SALINE) SALT DOME

Large deposits of rock salt are found in many countries throughout the world Some of the more important are in Austria Poland Czechoslovakia Germany Spain Great Britain and U S S R In the

not attainable in laboratory experiments at ordinary pressure and temperature but they resemble in many ways the characteristics of ductile metals deformed at high temperature. The mechanisms of flow and fracture of rocks with particular reference to the problems of inferring the processes of deformation of rocks in the earth are treated in this article. For a discussion of the aspects of rock mechanics related to building and the engineering applications of rocks *see* ENGINEERING GEOLOGY

Deformational behavior Rocks exhibit a spectrum of deformational behavior from brittle fracture through shear fracture and cataclastic flow to ductile flow by intracrystalline glide or recrystallization. In nature this spectrum is shown in any one rock type from brittle behavior when the rock is deformed near the surface at low pressure and temperature to flow when deformed at depths where pressure and temperature are high. In the laboratory any one rock type which is brittle under ordinary conditions may be made to pass through this spectrum of behavior as confining pressure temperature or both are increased. This transition from brittle to ductile behavior depends also on the nature of the stress system on the pressure of interstitial fluids and probably on the rate of strain although the latter dependence has not yet been explored.

H C Heard reports an experimental study of the brittle-to ductile transition in Solenhofen limestone as a function of all these variables except rate of strain. In dry compressive tests the transition occurs at 1 0 kilobars (kb) confining pressure at room temperature and the required confining pressure decreases nearly linearly to zero at 480°C. In dry extension tests however the confining pressure required for transitional behavior is 7 5 kb at 25°C, decreasing linearly to 0 7 kb at 700°C. In compressive tests with interstitial fluid pressure the confining pressure at the transition increases by 50% when the interstitial fluid pressure rises to 90% of the confining pressure. When the interstitial fluid pressure is equal to the confining pressure the transitional confining pressure is roughly tenfold that for dry specimens. Thus the depth in the earth required for flow must vary with the nature of the stress system and the magnitude of the pore pressure as well as with temperature.

Under sufficiently high confining pressure and temperature all rocks must become ductile deforming by uniform flow. In the laboratory rocks of the following types have been rendered ductile anhydrite limestone marble dolomite basalt dunite and pyroxenite. Quartz and quartzite are brittle under the highest pressures and temperatures yet tried (25 kb at 25°C. 5 kb at 800°C). Granite at 5 kb and 500–800°C exhibits a stress-strain relation similar to that of ductile metals but the deformation takes place by cataclasis in a thin zone of shear.

Recrystallization Rocks recrystallize in a manner similar to that of metals at high temperature. M E Buerger and E Washken demonstrated this with the minerals anhydrite, fluorite, periclase and corundum. The energy which drives this recrystallization is predominantly the strain energy in the deformed crystals.

Recrystallization has recently been produced in calcite one of the most commonly recrystallized minerals in nature. D T Griggs F J Turner and H C. Heard showed that marble undergoes syntectonic recrystallization while being deformed at temperatures from 400–800°C. The amount of recrystallization increases with increasing strain. This recrystallization reached a maximum at about 600°C for ordinary laboratory rates of strain and the temperature for maximum syntectonic recrystallization probably decreases as the strain rate decreases so that naturally deformed limestone probably recrystallizes at temperatures considerably below 600°C. Calcite recrystallized during deformation exhibits a very high degree of preferred orientation.

Annealing recrystallization in which a material is deformed cold and then heated in the absence of external stresses has also been produced in calcite crystals and aggregates. Griggs M S Paterson Heard and Turner report that calcite deformed cold will recrystallize when heated to a temperature of 500°C or higher. This recrystallization in the absence of an external stress field reduces the preferred orientation in deformed calcite aggregates tending toward a random orientation. In all these experiments on calcite, the solvents water and carbon dioxide have been shown to have no effect on the recrystallization.

In his study of experimentally deformed rocks Turner discovered the existence of internally rotated lamellae in deformed crystals. From the crystallographic orientation of these relict structures it is possible to infer the strain history of an individual grain and because deformation is largely homogeneous the bulk strain of the rock commonly can be deduced. This method checks the strains measured in the experiment to about 10%. It has been found by Turner and others that these internally rotated lamellae frequently occur in nature. It has been shown in calcite and dolomite rocks that these naturally produced rotated lamellae were formed by the same glide mechanisms that have been found in the laboratory. Thus a new tool for deciphering the tectonic history of rocks has been found.

Strength Strength at elevated temperature and pressures has been studied for a variety of rocks by J Handin Griggs and their coworkers. Under 5 kb (5000 bars) confining pressure (equivalent to that at a depth of 20 km in the earth) granite basalt pyroxenite and dunite all have a compressive strength of about 20 kb at 25°C. At 500°C, their strength has dropped to 10 kb. At 800°C, the strength of basalt is only 2 5 kb but the others are about 7 kb. Dolomite is 12 kb at 25°C. 10 kb at 450°C, and 6 5 kb at 800°C. Marble drops from

Rocket engine

A rocket eng ne provides the power to drive a vehi cle In rocket propulsion a reaction force is im

con ts of chemical compounds which when reacting with each other release energy thus differentiating this from other types of rocket engines which use nuclear or electric energy as their power source (see ELECTROMAGNETIC PROPULSION NUCLEAR ROCKET)

In this article are described the performance criteria, the basic features, the application and the preparation of various rocket engines which are driven by the combustion energy of chemicals Those rocket engines using liquid propellant and solid propellant will be described in some detail

Performance of rocket engines The performance of a missile or space vehicle propelled by a rocket engine is usually expressed in terms of such parameters as range, maximum velocity of flight maximum altitude or time to reach a given target Engine performance parameters (such as exhaust velocity specific impulse thrust or engine weight) are used in computing these vehicle performance criteria Table 1 gives typical performance values A change in rocket engine performance will often significantly affect the performance of the vehicle that the engine drives.

The momentum imparted to the vehicle by a rocket depends directly on the velocity v at which the hot gases are expelled from the rocket in a supersonic nozzle. This supersonic exhaust velocity is a good indicator of the effectiveness of any one propellant combination It is defined as

$$v = \sqrt{\frac{2gkR}{k-1}\frac{T}{M}\left[1 - \left(\frac{p_2}{p_1}\right)^{(k-1)/k}\right]}$$

where R is the universal gas constant k is the specific heat ratio of the reaction product gases, and g is the acceleration of gravity (32.2 ft/sec²) The pressure of the nozzle exit at optimum nozzle expansion area ratio is p_2 and the chamber pressure is p_1 Because of this dependence on the pressure ratio there is a slight increase (10 20%) in exhaust velocity as the exit pressure is decreased (increase in altitude) or as the chamber pressure is increased The propellants burn at very high combustion temperatures as shown in Table 2. The exhaust velocity increases as the molecular weight of the combustion gases M decreases and the combustion temperature T increases but exhaust velocity also increases if the pressure ratio across the exhaust nozzle increases

Specific impulse is another way of expressing exhaust velocity it is the exhaust velocity divided by g Its units are pounds of thrust per pound of propellant flow per second A high value is desired (Tables 1 2, and 3 and Fig 1) Specific impulse also increases with combustion temperature and decreases with molecular weight of the exhaust gases Exhaust velocities and specific impulses can be calculated from thermochemical relations the

Table 1 *Typical performance values of rocket engines**

Engine parameter	Typical values
Specific impulse at sea level	180–375 sec
Specific impulse at altitude	215–449 sec
Exhaust velocity at sea level	5800–12 000 ft/sec
Combustion temperature	4000–7500°F
Chamber pressures	1000–1500 psi
Ratio of thrust to engine weight	20–150
Thrust	1 2,000 000 lb
Flight speeds	0–40 000 ft/sec

*Exact values depend on application, engine design, and propellant selection

Table 2 *Typical performance of four categories of liquid propellants*

Propellant	Theoretical thrust chamber *specific impulse at 500 psi* chamber pressure sec		Bulk density lb/ft³	Optimum mixture ratio (oxidizer fuel)	Combustion temperature °F	Molecular weight of exhaust gas lb/mole
	At sea level area ratio = 8	In vacuum area ratio = 25				
Cryogenic						
Oxygen and kerosine	261	324	63	2 25	5800	22
Oxygen and 92 5% ethyl alcohol			61	1 5	5370	23
High energy						
Fluorine and hydrogen	364	447	19	4 0	4700	9
Fluorine and hydrazine	303	372	80	1 75	7300	18
Storable						
Nitric acid and dimethyl hydrazine	246	304	76	2 4	5100	22
90% hydrogen peroxide and kerosine			79	7 0	4600	22
Monopropellant						
90% hydrogen peroxide and hydrazine	137	167	87		1365 1800	21

United States bedded salt deposits underlie large areas of Kansas, Michigan, New York, Ohio, Oklahoma, Pennsylvania, and Texas More than 100 salt domes have been located in Texas and Louisiana

The use of salt is older than recorded history Because of the necessity of salt in the human body, its first use was undoubtedly as a food It was early an article of commerce and over many of the ancient trade routes salt was carried as a principal commodity Today salt finds its chief use in the chemical industry as the raw material used in the manufacture of hundreds of compounds of sodium and chlorine It is also used in meat packing, fish curing, refrigeration, and livestock feed

Salt is recovered from brines in many parts of the world by solar evaporation Sea water or saline lake water is introduced into shallow artificial basins where evaporation brings about concentration and subsequent precipitation Rock salt is also mined either in the conventional manner of sinking a shaft to the deposit or by dissolving the salt in water introduced into the salt bed and pumping the brine to the surface *See* CHLORINE, SALT (FOOD), SODIUM [C S HU]

Rock shell

Any of several species of the family Muricidae, class Gastropoda, phylum Mollusca Most authorities place the rock shells in the genus *Murex*, but others divide the group into several genera Not all the *Murex* are known as rock shells

These animals are found in all seas, but are most abundant in the tropics They are active, carnivorous snails, generally preferring gravelly or rocky bottoms Their shells are thick and solid, and usually spiny with three growth ridges to a whorl

The red rock shell, *Murex recurvirostris rubidus*, is a fairly common shell of the Florida coast It is about 2 in long and is mottled gray, brown, or pink in color, with a yellow operculum One of the more common Pacific species is *Murex* (*Ocenebra*) *interfossa*, the sculptured rock shell, which is common from Alaska to Lower California It is a small rock dweller, only about an inch long

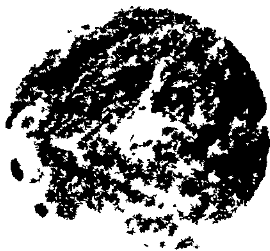
Murex (*Pterynotus*) *festivus*, the festive rock shell, may be 3 in or more in length It is common from Santa Barbara Calif, southward It is a dingy, rough shell, shading from white to gray, with rather elaborate ridges frills and sculpturings The young are sometimes brilliant scarlet This snail lives on weeds and trash, or sometimes on muddy rocks *See* GASTROPODA, SNAIL. [J D B]

Rocket

Either a propulsion system or a complete vehicle driven by such a propulsive engine. A rocket engine provides the means whereby chemical matter is burned to release the energy stored in it and the energy is expended, in this example, by ejection at high velocity of the products of combustion (the working fluid) The ejection imparts motion to the vehicle in a direction opposite to that of the ejected matter A rocket vehicle is propelled by rocket reaction and includes all components necessary for such propulsion, a payload such as an explosive charge, scientific instruments, or a human crew A rocket vehicle also includes guidance and control equipment mounted in a structural airframe or spaceframe [G P S.]

Rocket astronomy

The technique of obtaining astronomical data by means of instrument carrying rockets. Earth's atmosphere is virtually opaque to electromagnetic radiation of wavelengths outside bandwidths ("windows") ranging from 0 3 to a few microns and from 9 to 14 microns However, instruments transported beyond Earth's atmosphere by means of rockets can register radiation in spectral regions that are ordinarily absorbed by atmospheric gases



First complete ultraviolet photograph of the Sun Photograph taken at altitude of 123 miles by camera recovered from nose cone of Aerobee rocket launched March 13, 1959 at White Sands Experiment conducted by scientists of Naval Research Laboratory under auspices of the International Geophysical Year program (*Official U S Navy photograph*)

Instrument readings are usually transmitted by radio from rocket to Earth (*see* TELEMETERING) Certain types of information (for example, photographs) cannot be satisfactorily transmitted, and therefore arrangements are often made to protect and recover the pertinent equipment (see illustration)

Astronomical instruments have also been carried to high altitudes by balloons, and instrumentation to be carried by artificial satellites is under development.

Rocket astronomy first used in 1945 in the United States with German V 2 rockets, has been especially fruitful in studies of solar phenomena *See* ASTRONOMICAL SPECTROSCOPY RADIO ASTRONOMY, SUN [K.W F]

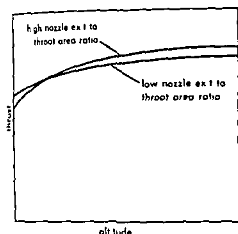


Fig 2 Typical variation of thrust with altitude

where m is the mass flow through the nozzle, v is velocity A_e is nozzle exit area and pressure difference $(p_2 - p_3)$ is that between nozzle exit pressure p_2 and ambient pressure p_3 As the mass flow increases so does the thrust Because p_2 is usually close to p_3 the second term of the thrust equation is small. As the altitude increases (p_3 decreases) the velocity and the thrust both increase as shown in Fig. 2. As exit area A_e is changed p_2 is changed also and this will affect the thrust. Since the pressure term is small the thrust is roughly equal to mass flow m and exhaust velocity v

Rockets generally have a high nozzle area ratio This is the ratio of the exit to the throat area it is usually between 20 and 60 for very high altitude

used to make engine weight as low as possible

... of it can be found from the integral of the thrust with respect to time

Mixture ratio is the ratio between the oxidizer mass flow rate and the fuel mass flow rate of a liquid propellant.

Total impulse-weight ratio is an indication of the over all design effectiveness of a rocket It is the total impulse divided by the total propulsion system weight, including tanks and propellants For pressurized liquid propellant rockets it has values between 80 and 130, for solid propellant rocket units, between 100 and 180 and for pump-fed liquid propellant units between 130 and 215

Liquid propellant engines Liquid propellants usually consist of a separate oxidizer for example liquid oxygen or liquid fluorine, and a separate fuel, for example gasoline, alcohol, or hydrazine (see PROPELLANT)

... they are injected mixed and burned at high pressures and very high temperatures to form the gaseous reaction products, which in turn are accelerated and ejected at high velocities The feed system for transferring the propellants into the thrust chamber includes valves and controls

With cooled thrust chambers it is possible to operate the liquid propellant rocket engine for extended durations Uncooled thrust chambers are suitable for use with propellant combinations having low combustion temperatures or those which operate for relatively short periods of time so that the uncooled walls will act as a heat sink and not overheat to the point that the material will no longer fulfill its intended function of containing the pressure

The principal components of a thrust chamber (Figs 3 and 4) are the nozzle, the injector, the chamber and in certain cases, the igniter A desirable rocket thrust chamber combines light weight construction with high performance, efficient combustion simplicity of design and fabrication satisfactory heat transfer from the hot gases to the walls and reliability The nozzle is usually of the converging diverging De Laval type, it accelerates the gases to high exhaust velocities The size and proportions of the nozzle have a critical influence on the chamber pressure thrust, propellant flow exhaust velocity of the thrust chamber, and the variation of these parameters with altitude and with each other A special contoured shape of the

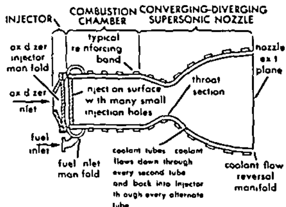


Fig 3 Cooled liquid propellant thrust chamber with contoured nozzle divergence section

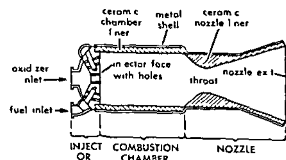


Fig 4 Uncooled liquid propellant thrust chamber with ceramic liners

Table 3 Characteristics of three types of solid propellants

Characteristics	Propellant type		
	Composite	Composite	Double base
Typical oxidizer %	NH_4ClO_4 50–80	NH_4NO_3, 80	Nitroglycerine, 30–45
Typical fuel %	C_2H_4O, 48–20	C_2H_4O 18	Nitrocellulose, 45–55 Other, 0–20
Combustion temperature, °F at 1000 psi	3000–5000	2400–2800	3000 5000
Typical specific impulse at 1000 psi sec exhausting to sea level	175–245	180–200	180–220
Burning rate in /sec at approximately 70°F and 1000 psi	0 1–0 7	0 1	0 2–0 8

values so obtained are known as theoretical specific impulses and theoretical nozzle exhaust velocities Actual values are 92–96% of their theoretical values since the losses due to friction, divergence, and incomplete combustion are relatively small

Thrust F is defined in terms of a momentum and a pressure term

$$F = mv + (p_2 - p_3) A_e$$

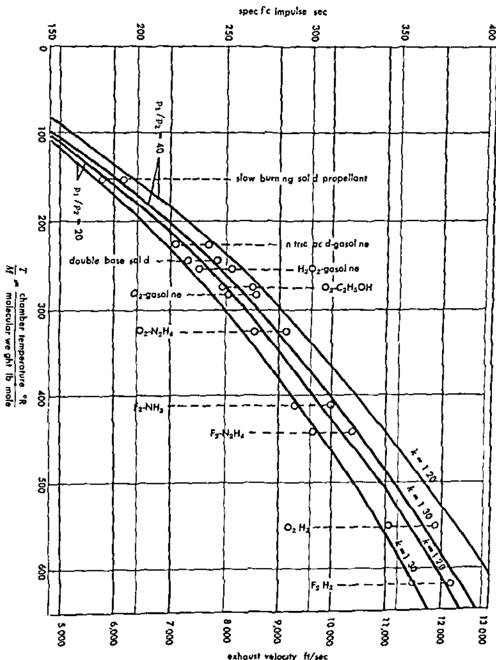
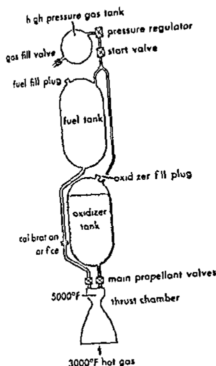


Fig 1 Specific impulse and exhaust velocity as a function of the chamber temperature and the molecular weight for various values of specific heat ratio k and nozzle pressure ratio Points for several propellant combinations are given



Fg 6. S mpl fied schematic of liquid propellant rocket eng ne with gas pressurized feed system

This is achieved in a chemical combustion chamber similar to the main thrust chamber but the propellants are burned at a different mixture ratio The assembly of pumps and turbine is designated a turbopump Like other rotating machinery it has bearings, seals high speed shafts, and lubrication devices. See COMPRESSOR

Turbopump feeds are generally superior to and

trifugal pump is generally considered superior to other types of pumps and is economical in weight

tant that the pressure at the inlet of the pump always be sufficiently high to avoid low pressure

ty of the rocket.

In addition to deriving thrust from a rocket engine, it is possible to obtain several auxiliary uses In a turbopump feed, for example it is fairly easy to obtain shaft energy for driving various auxiliaries in the missile, for example hydraulic, pneumatic, or electric generators The heat in the rejected turbine exhaust gases is often used for heating various fluids in a missile for example, the gases which are used for pressurizing tanks

Controls and operation All liquid propellant rocket engines have controls, which accomplish ... the following functions: start and

gency shutdowns when safety devices ... malfunction fill and drain propellants, and permit functional checkout of critical components without actual rocket engine operation

Starting stopping, and restarting are generally accomplished by valves in the main propellant lines Because of the high pressures and the large flows involved, these valves are often operated from separate pilot valves which control the energy going to these main valves

Rocket engine operations are programmed by automatic controls that set and hold the flow, the pressure, or the mixture ratio to the desired value for a given flight program Automatic thrust and pressure chamber controls program a specific ballistic missile flight in such a manner that successive flights can be made in a reproducible fashion A hydraulic, pneumatic, or electrical system is often used to actuate the principal valves to the thrust chamber and the gas generator and for furnishing power to any automatic controls and to the re

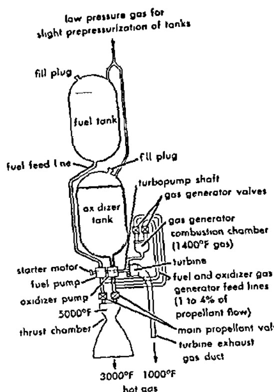


Fig 7 Simplified schematic of liquid propellant ro engine with turbopump feed system

divergent section permits a slight increase in performance compared to a conical nozzle

In the combustion chamber the burning of the liquid propellant takes place at high pressure, usually between 300 and 1000 psi An injector introduces and meters the flow of the liquid propellants and also atomizes and mixes them in the correct proportions in such a manner that they can be readily vaporized and burned (Fig 5)

Certain types of fuels and oxidizers start burning when they come in contact with each other. These self reacting combinations, such as hydrogen peroxide and hydrazine are known as hypergolic propellants Other types of propellants for example, oxygen and alcohol, are not spontaneously ignitable, but require an igniter to furnish thermal energy for starting their combustion reaction Electrical energy supplied by sparks or hot glow plugs pyrotechnic powder squibs of short duration or the injection of hypergolic chemicals initiate the combustion In some cases, particularly in variable thrust rocket engines a small precombustion chamber is built into the injector to maintain a flame throughout the variety of altitude and thrust conditions over which the thrust chamber must operate and ignite

Cooled thrust chambers permit one or the other of the two propellants to circulate through a cooling jacket which surrounds the combustion chamber and the nozzle, this prevents the wall material

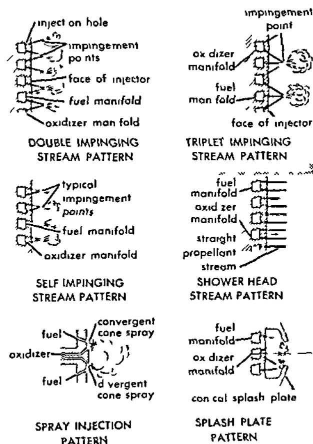


Fig 5 Schematics of several types of injectors (*From G P Sutton, Rocket Propulsion Elements, 2d ed , Wiley, 1956*)

from overheating and permits operation for long durations The use of one of the propellants as a coolant is often termed regenerative cooling because of its similarity to the regenerative steam power cycle Film cooling is a method whereby a thin liquid film covers and protects the exposed wall from excessive heat transfer The film is introduced by injecting small quantities of one of the propellants at low velocities in a number of places along critical exposed surfaces Sweat cooling or transpiration cooling uses a porous wall material which admits a coolant uniformly over a surface

Since cooling permits these chambers to operate at relatively nominal wall temperatures, various common metals are suitable for construction Uncooled thrust chambers are generally made of materials with a high heat capacity, such as copper or steel A protective layer of ceramic or other special coating is used to reduce the amount of heat that needs to be absorbed by the metal, this construction permits longer duration of operation Uncooled thrust chambers with durations of over a minute have been used where combustion gas temperatures are low

Feed system The feed system of a liquid propellant rocket engine transfers the liquid propellants from the vehicle storage tanks to the thrust chamber, it has a power source, which furnishes the required transfer energy, and control devices for regulating the rocket propellant flow and therefore the engine performance

The gas pressure feed system offers one of the simplest and most common means of transferring propellant and oxidizer by displacing them with a high pressure gas which is fed into the tanks under a regulated pressure (Fig 6) The stored high pressure gas furnishes the pressurization energy The thrust of a pressurized gas rocket propulsion system is determined by the magnitude of the propellant flow which is controlled by the gas pressure regulator setting For low thrust and short duration this feed system is generally lighter and superior to other more complicated ones In some single-use applications such as an expendable air launched missile a simple pressure feed system that may use burst diaphrams instead of valves is satisfactory When repeated use is desired such as in aircraft assisted take off units a special thrust regulating device a tank level gage fill and drain provisions, and other components and functions are added The application dictates the complexity of the feed system In some cases a chemical gas pressurization is substituted for the stored inert gas this arrangement saves weight but sacrifices simplicity

In a turbopump feed system the propellant is pressurized by means of pumps driven by one or more turbines (Fig 7) which derive their power from the expansion of hot gases A separate gas generator ordinarily produces these gases in the required quantities and at a temperature which will not hurt the turbine buckets (1200–1800°F)

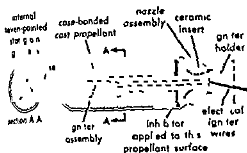


Fig 9 Solid propellant rocket with case-bonded internally burning grain

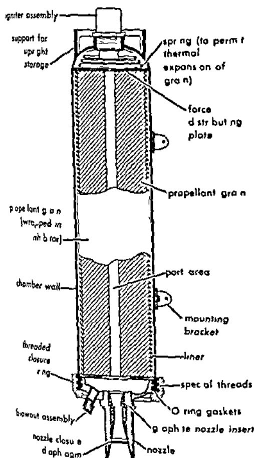


Fig. 10 Lined spring held solid propellant grain for assisted take off rocket unit with blow-out device (From G P Sutton, Rocket Propulsion Elements 2d ed Wiley 1956)

ally made of high strength material and is often protected from excessive heating by internal insulating layers made of either a very slow burning propellant having a low burning temperature or a ceramic insulator The grain is often pasted to or cast directly into the chamber case so that the grain forms its own heat barrier to prevent over heating of the chamber walls This is known as case bonding (Fig 9) The heating of the inner wall surfaces produces a temperature gradient across the wall and thereby induces thermal stresses in the wall Insulation minimizes this. Combustion chambers are usually cylindrical in shape with elliptical or spherical ends

One end of the chamber is usually detachable to permit the assembly or loading of the grain Burst diaphragms or other safety provisions to prevent over pressurization of the chamber are sometimes included In some designs the burst diaphragms are intentionally activated to stop the application of thrust at a predetermined time The chamber has means for holding the grain in place and positioning it so that it will not be forced into the nozzle due to the vehicle's acceleration In addition the chamber usually has provisions for mounting to the vehicle for allowing some differential thermal expansion between the metal parts and the grain and for sealing against moisture which would cause deterioration of certain grain chemicals. Careful design of the thrust chamber produces a lightweight unit

The most severe heating of the hardware takes place at the exhaust nozzle Here the high velocity gas at the high combustion temperature will oxidize soften wear and erode the nozzle material unevenly For this reason it is often desirable to put special heat resistant ceramic or graphite inserts into the nozzle throat region to minimize unsymmetrical enlargement of the nozzle area Ceramic nozzles have been successful for long periods of time however for high temperature propellants their effectiveness seems to be limited to less than a minute With low temperature propellants durations in excess of 10 minutes may be feasible Metal nozzles with protective coatings have been successful for short durations Extra metal thickness is usually added to absorb additional heat An accurate alignment of the nozzle axis with the center of gravity of the flying vehicle is essential to minimize flight errors particularly in unguided vehicles

The igniters used for solid propellant rocket units are almost exclusively of the pyrotechnic type A pyrotechnic igniter usually consists of an electrically heated wire surrounded by a small amount of sensitive primer powder charge which in turn ignites a larger main igniter charge In some cases the igniter is put into the forward end of the chamber so that ignition gases sweep past the complete propellant charge before reaching the nozzle

Grain shape and burning The thrust of the rocket is approximately equal to the product of the mass flow and the effective exhaust velocity The mass flow rate is equal to the product of the exposed burning area the rate of burning and the density of the propellant By designing the grain so that more or less surface is exposed the designer may increase or decrease the thrust (Fig 11) It is also possible to limit the exposed burning surfaces by applying inhibitors which are special inert or slow burning chemicals Grains that have much inhibitor applied to their exposed surfaces called restricted burning rockets are differentiated from grains that burn on all exposed surfaces

mainder of the vehicle system Often the fuel pressure itself can be used to actuate some of the controls

One special form of control, known as the propellant utilization system is an automatic device that insures that both the fuel and the oxidizer tank are emptied simultaneously This is achieved by continuously measuring the amount of propellant that remains in both of these tanks and by adjusting throttling valves in both of the main propellant lines leading to the thrust chamber

The control of the thrust vector attitude can be achieved in a number of different ways (Fig 8)

In the operation of liquid propellant rocket engines it is often desirable to vary the thrust or to throttle the rocket engine In multiple rocket engines this can be achieved by shutting down some of the thrust chambers while letting others continue to operate In a single thrust chamber rocket engine, the throttling can be achieved by reducing the chamber pressure and the flow

Propellants Liquid propellants include, besides the oxidizer and the fuel catalytic compounds, which combine oxidizer and fuel properties in a single substance A bipropellant rocket unit has two separate propellants a fuel and an oxidizer, which are not mixed until they come in contact with each other in the combustion chamber Most liquid propellant rockets have been of the bipropellant type A monopropellant contains oxidizing agent and combustible matter in a single substance It can be a mixture of compounds, such as hydrogen peroxide with liquid alcohol or it may be a homogeneous chemical agent such as nitromethane Characteristics and properties of the propellants and the resulting gaseous reaction products determine the choice and selection of propellants for specific applications Economic and performance parameters, physical and chemical properties desirable for storage and operation, as well as consideration of common physical hazards are considered Typical values for propellants are given in Table 2

Solid propellant rocket engines. In a solid propellant rocket engine the propellant is contained in solid form within the combustion chamber The hardware includes, in addition to the combustion chamber, a nozzle, an igniter, and provisions for mounting the rocket (Figs 9 and 10) Solid propellants themselves usually have a plastic, cakelike appearance and burn on their exposed surfaces to form hot exhaust gases which are ejected through the nozzle The physical mass or body of the propellant is called the grain In some rockets there is more than one grain inside the same combustion chamber

Engine components The solid propellant grain contains all the material necessary for sustaining combustion It can be a mixture of several chemicals, for example a mixture of ammonium perchlorate in a matrix of organic fuel such as rubber, or it can be a homogeneous charge of special oxidizing organic chemicals such as nitrocellulose or nitroglycerine Once the propellant is ignited the grain burns smoothly on all of its exposed surfaces in a direction normal to the burning surface

Compared to liquid propellant the solid propellant requires no pump or pressurization for the fuel tank and hence is mechanically simpler The combustion chamber with solid propellant is larger especially for large rockets and is frequently operated at higher pressure than for a liquid engine Solids are more storable and are more immediately ready for use

The chamber confines the propellant grain and the reaction products of combustion It is gener

method	gimbal or hinge	jet vanes	small control thrust chambers	gas bleed from main thrust chamber	turbine exhaust gas control	jetevator
description	universal joint	four rotating heat resistant aerodynamic vanes	two-gimballed small auxiliary thrust chamber	two or four cooled bleed passages with valves	gimbal on turbine exhaust nozzle	gimbal suspended rotating collar near nozzle exit
loss in specific impulse	negligible	2 5%	1 10%	1 7%	1 2%	2 6%
limitation	pitch and yaw control only	← good for pitch yaw and roll control →				pitch and yaw only

Fig 8 Devices for obtaining thrust direction control from rocket engine

in the end of the thrust chamber opposite that of the nozzle so that the majority of the gases go in a direction that produces no useful thrust

Most solid propellant grains are somewhat sensitive to variations in the ambient temperature and have a tendency to become soft on very hot days or brittle on very cold days for certain propellants it is therefore necessary to restrict the temperature range over which they may operate If the propellant should crack from too great a temperature variation, which induces thermal stresses or from rough handling additional burning surfaces would be created in the cracks and an unregulated increase in the pressure would cause failure of the chamber Care is necessary in handling propellants and solid rocket units gently so as to not subject them to sudden shocks by dropping or jarring

There is a maximum pressure above which smooth combustion can no longer be sustained and detonations may occur, and a minimum pressure below which stable consistently smooth burning does not seem possible Because of the active nature of the chemicals some propellants deteriorate in storage, this can often be prevented by the addi

Table 4. Application of rocket engines in guided missiles

Engine purpose	Thrust lb Duration sec Acceleration g	Propellants and specific impulse (I_s) sec	Engine type and special features
Assisted take-off shortens ground run or increases take-off load	250–10 000 lb 10–50 sec Up to 0.5 g	Solid I_s = 190–210* Storable liquid (nitric acid dimethylhydrazine) I_s = 200–250*	Expendable droppable refueled and reused for next take-off some have canted nozzles
Superperformance additional thrust and airplane performance for fighter aircraft at high altitude	2000–10 000 lb 100–500 sec *1.0 g in any direction* during maneuvers	Storable liquid (H_2O_2–N_2H_4) I_s = 235–260†	Repeated starts throttled thrust for climb and *cruise improve rate of* climb of aircraft higher flight ceiling altitude
Principal aircraft drives research aircraft to and at high altitude and speed	20 000–90 000 lb 60–180 sec at full thrust, 120–600 sec additional at partial thrust *0.5–4.0 g in any direction*	Storable liquid I_s = 235–260† Cryogenic liquid I_s = 280–310† *High-energy liquid* I_s = 340–400†	Same as superperformance engine gimbal for attitude control
Air-to-air guided missile accelerates missile	1000–30 000 lb 0.3–10 sec 2.0–100 g (also high side accelerations)	Mostly solid a few storable liquids I_s = 200–250†	Simple single-shot device liquid uses prepackaged pressurized system may have stepped thrust for cruise operates under *wide temperature limits*
Air-to-surface boosts and sustains missile	1000–30 000 lb 0.4–100 sec 2.0–100 g (also high side accelerations)	Mostly solid some storable liquid I_s = 190–240	Same as air-to-air guided missile
Surface-to-air (usually 2- or 3-step missile) antiaircraft or antimissile missile	*10 000–100 000 lb* (booster stage) 1000–15 000 lb (sustainer stages) 1.0–30 sec (booster) 10–180 sec (sustainer) 2.0–100 g (also high side *accelerations*)	Usually solid propellant sometimes storable liquid I_s = 190–250	Simple engine with possible thrust programming
Surface-to-surface medium and long range ballistic and winged missiles of 50 to 8000-mile range	10 000–400 000 lb (booster) 3000–100 000 lb (sustainer) *30–300* sec 4–10 g	Cryogenic liquid (oxygen kerosine) I_s = 220–265* Solid I_s = 210–245*	Turbopump-fed high *performance liquid* engine, attitude control by gimbal high energy solid with thrust termination device attitude *control* by jetvanes or jetevator accurate thrust cutoff
Surface-to-surface short-range infantry support and antitank missile	*50–20 000* lb 0.2–5 sec Up to 100 g	Solid I_s = 170–230*	Simple prepackaged operates under wide temperature limits simple launcher

* At sea level. † At altitude

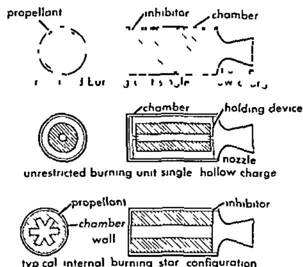


Fig 11 Typical solid propellant grain configurations (From G P Sutton, Rocket Propulsion Elements, 2d ed, Wiley, 1956)

Holes that are put into the grain to increase the burning surface are known as perforations

If the grain design is such that the thrust increases during the operation of the rocket, it is called progressive burning, if thrust decreases during operation, it is regressive burning

The burning rate of solid propellant is measured in a direction normal to the burning surface and is expressed as a linear distance per unit time It usually varies as the chamber pressure to a power between 0 4 and 1 0, and is also a function of the initial ambient temperature The same grain gives more thrust but a proportionally shorter firing duration at higher temperatures (Fig 12) The chamber pressure can be varied by changing the throat area of the nozzle or the exposed grain burning surface

It is possible to vary the thrust of a solid propellant unit not only by judicious design of the grain so that the burning surface varies with duration, but also by using different kinds of propellants, each with a different burning rate, to be exposed as the initial layers are consumed

Solid propellants The solid propellants themselves are the chemicals that constitute the grain and that produce hot high pressure gases in a combustion process

Early rockets used black powder (a mixture of nitrate, sulfur, and a carbonaceous material) which is a loose granular powder The use of tamped granular grains is generally unsatisfactory for high performance applications A solid propellant usually contains (1) oxidizers such as nitrates or perchlorates, (2) fuels such as organic resins, plastics, or rubbers, (3) a chemical compound combining both fuel and oxidizer qualities such as nitrocellulose or nitroglycerine, (4) additives to control the fabrication process and such grain properties as burning rate, physical properties, and chemical deterioration, and (5) inhibitors, which are bonded on to the propellant to restrict the burning surface

Of the many varieties of propellants, two are significant: composite propellants and double base propellants The former consist of crystalline, finely ground oxidizer particles dispersed in a matrix of a fuel compound, the latter contain largely unstable chemical compounds such as nitroglycerine or nitrocellulose which are capable of combustion of themselves in the absence of other materials Because these propellants are often based on the above two compounds they are often known as double base propellants Different propellants together with their properties are listed in Table 3 In general, the fabrication is comparable to the manufacture of complex refinery products, plastics, or rubber

Solid propellants are frequently cast and case bonded to the combustion chamber Casting requires a fuel binder mixture that can be safely handled as a liquid and that cures or solidifies after it is poured into the case Case bonding places stringent requirements on the mechanical properties of the propellant About 65–85% of the propellant is dispersed powder, the mechanical properties being controlled by the remaining 15–35% of the matrix This binder must absorb a wide range of stresses and differences in thermal expansion

Double base propellants are generally extruded and are not case bonded, thus the mechanical requirements on the grain are less stringent However, the chronologically older double base propellants are being replaced by cast, case-bonded composite propellants that permit larger grains of higher specific impulse

Operation Solid propellant rockets are started by activating the igniter They normally burn until all the grain is consumed It is possible in some cases to add provisions to stop the burning at a predetermined time by means of burst diaphrams which open additional nozzle area and lower the pressure in the combustion chamber below the value at which steady combustion can be sustained A similar effect can be achieved by opening a hole

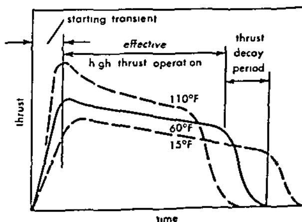


Fig 12 Thrust time variation of a typical solid propellant rocket with slightly regressive burning characteristics. Dotted alternate curves show effect of temperature

reinforced buildings of heavy construction that can withstand a blast, an extensive series of personnel warning devices special fire fighting equipment, limitations on the amount of propellant in any one area and on the distance between buildings the use of sparkproof tools and a rigorous training of personnel. A typical facility for testing large rocket engines in a vertical position is shown in Fig 13

The instrumentation is usually of the remote indicating or remote recording type Pressures temperatures, forces, operating sequence, flows, vibrations, and strains are some of the more common parameters measured Special instruments are needed for many of these, for example, a pressure

Storage, transportation, and servicing. As with other precise pieces of equipment, rocket engines must be handled carefully Special equipment and procedures are used for storing transporting and maintaining these engines Moisture-tight transport containers are used which also provide some insulation against shock For heavier engines special lifting and handling dollies are provided For

be maintained in instant readiness for long periods of time. [G P S]

Bibliography G P Sutton, *Rocket Propulsion* ... *Missile Propulsion*, vols 1 and 2 1958

Rocket staging

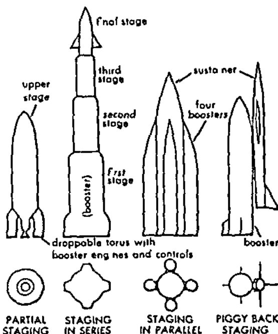
One way to minimize the weight of large missiles or space vehicles is to use multiple stages The

...on system, and has its own tankage and control system

Once the propellant of a given stage is expended the ...

succeeding stages By dropping off this useless weight, the mass that remains to be accelerated is made smaller, therefore it is possible to accelerate the payload to higher velocity than would be attainable if multiple staging were not used

The table shows that the use of multiple stages in ambitious space missions gives a larger improvement in the weight of passengers and equipment that can be carried than its use in missions with less range Various staging configurations can be used Multiple stage arrangements are used in large



Typical schemes for staging missiles

space vehicles in some long range ballistic missiles various research rockets and in certain anti aircraft and air launched military missiles

Ideally it would be desirable to stop the propulsion unit of the operating stage at the same time that the operation of the propulsion unit of the next stage is initiated This close timing is not possible There are inherent difficulties in the separation of the stages caused by drag forces the lack of gravity in space the noninstantaneous starting and stopping of rocket engines and the possible interference of various stages Often a positive mechanical separation mechanism is included in the design that forces the two stages to separate

By staging a missile it is possible to improve its performance either by providing more range, more altitude faster flight, or more payload or by reducing the gross weight and the size of the missile for a given specific mission without diminishing the payload It is quite possible to employ different types of power plants different types of propel

Total payload as a function of the number of stages for specific missions. *(Payload values are relative to each other)*

Mission	Missile take-off weight, lb	Payload lb, for no of stages			
		1	2	3	4
1500 mile ballistic missile	50 000	2000	2700	†	†
6000 mile ballistic missile	500 000	*	6000	9000	†
Moon impact	500 000	*	100	1000	3000

Table 5 Application of chemical rocket engines in space flight

Characteristics	Booster engine	Sustainer engine	Terminal stage engine
Purpose	Lift vehicle off ground	For intermediate vehicle stages	Final powered stage to give accurate final velocity to vehicle
Thrust lb	High up to 6 000 000	2 000–150 000	Low less than 2 000
Duration sec	50–150	120–400	30–1000
Propellant and specific impulse (I_s) sec (at altitude)	Cryogenic liquid (oxygen kerosine) $I_s \approx 300$–315 High-energy liquid (fluorine-hydrazine) I_s = 360–410 Solid I_s = 260–290	Same as booster	Same as booster
Engine type	Turbopump feed system with liquid propellant lightweight case for solid propellant	Turbopump or pressurized feed system with liquid propellant lightweight case for solid propellant	Requires accurate cutoff sometimes throttling or restart
Maximum acceleration g	4–10	1 10	Low less than 1
Special features	Attitude control by gimbal vernier jetevator	Same as booster sometimes operated during booster firing to augment take-off thrust altitude start required	Same as sustainer

as polyethelyne leads to a hybrid rocket. The solid fuel is arranged much like the grain of a solid propellant rocket so that it burns on its surfaces in the presence of the liquid or gasefied liquid oxidizer

Applications of rocket engines Tables 4 and 5 show the wide variety of applications of chemically propelled rocket engines in the field of aircraft space flight and missiles

Manufacture The hardware for rocket engines is fabricated by conventional techniques using mostly metallic materials Some of the weight limitations and the unusual high heat transfers require a few unusual fabrication processes and methods For example thin walled metal tubes used for the construction of cooled liquid propellant thrust chambers need special forming and joining processes The materials are usually conventional and include stainless steels aluminum alloys and high strength steels They have to be compatible with the chemical action of the propellant In some solid propellant rockets plastically bonded glass fibers have found acceptance because of their high strength weight ratio

Testing Because flights of rocket propelled vehicles are usually fairly expensive and because it is difficult to obtain sufficient and accurate data from fast moving flight vehicles it is accepted practice to test rocket engines and components extensively on the ground under simulated flight conditions Components such as an igniter or a turbine are tested separately Complete engines are tested in static engine test stands the complete vehicle is also tested statically In the latter two tests the engine and vehicle are tied down Only in flight tests are they allowed to leave the ground

The static test facilities must be capable of handling and disposing of the hot gases (5000°F) that are expelled from the rocket For those rockets that must be tested in a vertical position water cooled metallic flame buckets are used to prevent ground erosion

Since rocket propellants are dangerous to handle (they can explode with an energy higher than an equivalent weight of high explosive) special precautions and safety provisions must be observed in fabrication shipping and testing These include

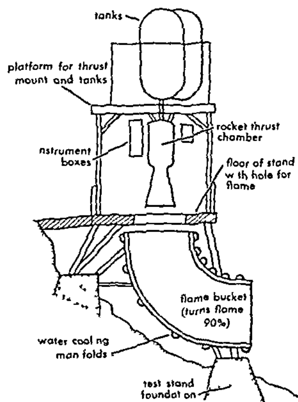


Fig 13 Rocket engine static test stand with elbow shaped jet deflection bucket

Principal features of four major tracks

	Elevation above sea level, ft	Track length, ft	Gage, in	Size, lb/yd	Rail length, ft	Continuous or welded
Edwards[a]	2300	20 000	56 5	1 1	39	Yes
Holloman[b]	4000	35 000	84	1~1	39	Yes
Hurricane[c]	5100	12 000	56 5	10?	39	Yes
SNORT[d]	2100	21 550	56 5	171	50	No

[a] Air Force Flight Test Center Edwards Air Force Base California.
[b] Air Force Missile Development Center Holloman Air Force Base New Mexico
[c] Hurricane Supersonic Research Site Hurricane Mesa Utah
[d] U.S. Naval Ordnance Test Station, China Lake California

and prevent derailing They may be powered by either liquid or solid propellant rocket engines (Fig 1) In the interest of maximum operating economy test vehicles are designed for multiproject use whenever possible A test cab might be used for several seat ejection projects (Fig 2)

Rocket sleds are recovered primarily through use of a water braking system This system utilizes the principle of momentum transfer, and consists of a water source located between the rails and a scoop and water turning channel mounted on the test vehicle. The gradient of the track through the braking area is downward, therefore the scoop on the sled takes a continuously deeper bite into the

ward and to the sides

Time station history of the sled during a test is produced by a magnet in conjunction with track coils wired to a recorder

Track programs Tests conducted on the track provide initial design data developmental evaluation, and performance measurements of completed components and equipments.

Missile components Such missile components as radomes airfoils, engines and guidance and control equipments are tested Test conditions include rain for erosion measurement stable platform with variable acceleration flutter testing and impact into various media

Escape mechanism Air crew escape methods and installations are tested on rocket sled tracks for almost all experimental and operational military aircraft Test ejections and jettisonings are conducted at supersonic velocities Data are recorded on seat and canopy trajectories, accelerations and decelerations of seat and dummies, aerodynamic stability of seats and canopies noise levels at dummies' ears, and pressure levels at critical cockpit areas before and after canopy removal

Flutter testing Testing of full scale missile and aircraft empennage and airfoil sections is accomplished on the rocket sled track facility It is a highly satisfactory means of investigating flutter characteristics

Parachute recovery Tests are conducted to evaluate and develop basic parachute materials and shapes Complete parachute systems are also tested Several revolutionary changes in design concepts resulted from rocket sled tests

Rain erosion testing The all weather reliability of guided missiles and radar aircraft depends upon the ability of their radomes to withstand the effects of rain at supersonic speeds Tests on full scale radomes are conducted by carrying the test item through a simulated rainfall range at the desired speed and prescribed time interval

Structures and materials External force loading is accomplished on test items through the application of the desired changes in acceleration or deceleration and rate of onset Through programming

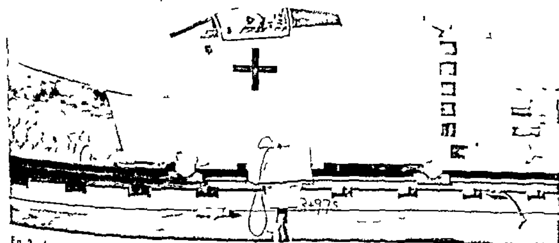


Fig 2 In tests of a pilot escape mechanism an anthropometric dummy simulates the pilot High speed camera at top rear of test vehicle records performance of ejected canopy dummy and seat Solid propellant rockets at rear of sled drive it High speed trackside camera in foreground provides additional data (Official U S Air Force photograph)

lants and entirely different configurations in successive stages of any one multistage missile Because staging results in additional complications to the missile it is generally impractical to have more than four or five stages in any one vehicle (*see* ROCKET ENGINE) [G P S]

Rocket sled testing

A method of subjecting structures to high accelerations or decelerations under controlled conditions The test object is mounted on a sled chassis running on precision steel rails and accelerated by rockets and decelerated by water scoops This captive track testing of full scale aircraft and missile components makes possible the recovery of expensive test sections facilitates the instrumention of the test and the reception and recording of the data and provides a degree of repeatability not normally achieved under conditions of free flight The components tested may be the rocket engines themselves sections of flight vehicles or equipments intended to operate under conditions of high acceleration or under other specific environments such as structural vibration acoustic vibration due to sound pressure and related aerodynamic conditions

Test facilities Sled facilities for captive flight testing have developed rapidly since their inception in 1948 Such facilities play a major role in research and development of aircraft and missile components and equipment Captive flight testing of complex and expensive structures is economically desirable In addition the recovery of the test specimen for post test examination greatly increases the information derived from the tests High speed track testing provides means for developing dynamic loads closely simulating those of free flight and the assurance that the test vehicle will follow a programmed trajectory on the test range enabling parameter measuring devices to record data on environment and performance with a precision otherwise unobtainable under dynamic conditions This type of facility permits the testing and evaluation of sophisticated problems associated with manned and unmanned missiles

Captive flight testing provides a high degree of repeatability of test conditions Within the track operational limitations specified velocities accelerations and rates of change of acceleration can be accurately repeated as many times as desired Tests are conducted on general purpose tracks at speeds in the vicinity of Mach 4

Four large general purpose tracks and twenty or so small special purpose tracks operate in the United States A summary of the principal physical features of the four general purpose tracks is presented in the accompanying table In general the test facility consists of a high speed rocket sled track of precision aligned crane or similar type rail fastened to a concrete foundation Supporting facilities available include rocket sleds to carry test specimens at required velocities vehicle recovery system instrumentation to sense and record the significant parameters complete photographic instrumentation and data reduction resources

Test sled The test vehicle takes many diversified forms ranging from the highly faired aerodynamically clean type used to explore the transonic and supersonic speed ranges to the relatively simple all purpose utility type used for testing in the subsonic and low transonic ranges All sleds are carried on shoes or slippers that grip the railhead

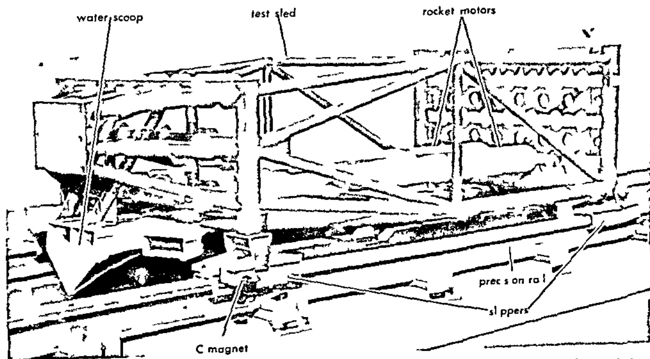
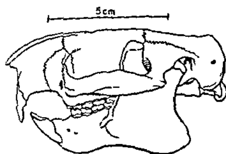


Fig 1 Pusher sled for multistage firings Fully loaded with 21 solid propellant rocket motors sled receives 200 000 lb of thrust (*Official U S Air Force photograph*)



Skull and jaw of *Paramys delicatus*, a middle Eocene rodent from Wyoming (After W Matthew, 1910)

incisors $\frac{1}{1}$ canines $\frac{0}{0}$ premolars $\frac{2}{1}$, molars $\frac{3}{3}$ *See* DENTITION

The oldest known rodents (late Paleocene) are squirrel like in appearance They seem to have replaced the multituberculates, the earlier gnawing mammals Ancestry of the rodents is not known but the order clearly belongs with the placentals By the end of the Eocene rodents had evolved into

peared in the Miocene eight in the Pliocene and two in the Pleistocene. There are only four families with no known fossil record *See* MULTITUBERCULATA.

From Eocene on rodents spread widely throughout North America, Asia and Africa. American hystricomorphs probably were transported across seaways on drifting vegetation from Central Amer

rodents probably dispersed to Australia in a similar manner sometime in the late Tertiary No other families reached Australia

The squirrels (Sciuridae) and cricetids (Cricetidae) eventually reached every continent except Australia. Other families like the beavers (Castoridae) and jumping mice' (Zapodidae) were restricted to the Northern Hemisphere, while others

evolution of morphologic characters [R A ST]

Rodenticide

A type of lethal chemical agent used to kill rodents, whether applied in bait, as a dust, or as a gas. The term also commonly refers to toxic materials used to kill a few mammals that are not rodents such as moles, rabbits, and hares

Rodents become a pest whenever man provides them with a favorable habitat (food and shelter), and, regardless of the density of predators, certain field rodents may become numerous whenever man's land use practices inadvertently create favorable homes for them

Anticoagulant rodenticides, such as diphacin, fumarin, PMP, pival, or warfarin sold under a variety of trade names effectively control rats and mice by producing internal bleeding that leads to death Small amounts of these anticoagulants must be consumed for several days in succession Be- [illegible] concentrations employed in baits

[illegible]

pets

Rodenticides generally recommended to the public for use on rats and mice are the anticoagulants mentioned above and less frequently, Antu [illegible] for Norway rats only), red [illegible]

and field rodents include the above [illegible] strychnine (either as the alkaloid or as the sulfate) Rodenticides now seldom used are arsenic trioxide barium carbonate and yellow phosphorus Poisonous gases commonly used are calcium cyanide $Ca(CN)_2$ which in the presence of moisture, forms hydrocyanic acid gas (HCN), carbon monoxide CO (automobile exhaust) carbon disulfide, CS_2 and methyl bromide CH_3Br

Rodenticides used as contact poisons in the form of dusts include Antu for Norway rats DDT for house mice and dieldrin and endrin for field rodents The last two are sometimes used as liquid sprays, as is also toxaphene Probably the most efficient rodenticides especially for field rodents are sodium fluoroacetate (Compound 1080) and thallium sulfate in that order These materials are not available to the general public because of their secondary poisoning hazard and the lack of good antidotes.

There is a continuous search for an ideal rodenticide and bait that is selective for the species concerned hence safe to other native fauna, humans and livestock easily counteracted with an antidote free of any secondary poisoning hazard to animals that might eat poisoned rodents something for which rodents cannot create a tolerance, painless odorless tasteless, and slow acting to minimize the development of bait shyness (poor reacceptance) by rodents No rodenticide is universally effective and both rodenticide and bait material should be periodically changed *See* PESTICIDE, RODENTIA [W E H]

Roentgen unit

A unit of ionizing radiation based on the energy absorbed in air when the air is exposed to radiation Because the changes induced in living tissue

of solid and liquid propulsion motors the desired accelerations and changes in accelerations are obtained If particular decelerations and changes in decelerations are desired the rocket sled is programmed to engage various restraint media and devices such as water braking friction braking air braking and cable restraint systems

Propulsion motors The usual thrust package on a rocket sled consists of a cluster of similar or mixed types of rockets Clusters of up to 20 medium sized rockets and clusters of up to 60 smaller rockets for specialized applications are used in a single firing Rocket configurations with total thrusts of 200 000–300 000 lb are not uncommon for a single firing

Varying thrust requirements often result in a need for firing the rockets in stages In motion firings are usually accomplished with static trackside power packages to hold sled weight to a minimum The condenser discharge technique uses a track mounted electrically charged screen which when ruptured by a sled borne conductor supplies the necessary electrical charge to the sled rocket firing circuit As many as 4 stages have been fired in a single test to control acceleration Coordinated firings of solid propellant rockets and liquid propellant rockets are sometimes made to obtain the advantages of each

Solid propellant rockets are frequently attached to basic liquid propellant rocket vehicles and fired concurrently They can be fired at the start of the test to increase initial acceleration near the end of the liquid propellant rocket firing as a boost during the test interval when air drag neutralization of an appreciable part of the available thrust can be disadvantageous or when the liquid propellant is expended to sustain a test velocity

Liquid propellant rocket pushers are sometimes used in conjunction with test vehicles which contain solid rockets for second and subsequent stages which ignite after shut down of the pusher vehicle separating the vehicles and leaving the pusher behind

Instrumentation Diversified instrumentation systems support the test programs Commercial and specially developed equipments are used Most tests require radio telemetry frequency modulation multiplex is used to provide up to 84 separate analog channels some with responses of a few cycles per second and others with responses up to 3 kc Occasionally tape recorders are carried on the sled

Motion and still photographs are regularly taken of tests Most commonly these are 16 mm motion pictures made at 24 frames per second (fps) to 25 000 fps in black and white Normal color exposures are limited to 1000 fps For greater resolution 70 mm pictures at 20 to 50 fps are taken or still pictures ranging from 35 mm through 4 by 5 in and larger

To provide space-time data metric ranges are installed parallel to the track Photographic data from such ranges record trajectories as long as 9000 ft and as high as 500 ft Sled borne cameras may supplement trackside equipment *See* AIRCRAFT TESTING WIND TUNNEL [E C.]

Rodentia

An order of mammals characterized by a single pair of enlarged chisel like incisor teeth The other incisors the canines and the anterior premolars are missing in both jaws leaving a characteristic gap between the chisel teeth and the cheek teeth This is much the largest and most complex order of eutherian mammals more than 350 genera of living rodents and nearly as many genera of fossil rodents are known There are more species of rodents than of all other mammals combined Rodents first appear in the fossil record in the Paleocene near the beginning of the Tertiary The oldest known fossil rodents have all the features of typical extant rodents and consequently the derivation of the order from more primitive mammals can only be inferred

The order Rodentia is divided into three great suborders The Sciuromorpha includes the squirrels and their relatives together with the mountain beavers true beavers kangaroo rats and pocket gophers These are the most primitive of living rodents many of them differing little from their Paleocene ancestors The Myomorpha includes the tremendous variety of rats and mice dormice and jerboas The Hystricomorpha includes the porcupines cavies agoutis and spiny rats The hystricomorphs are essentially a South American group

Throughout the Tertiary the rodents were the most numerous of all groups of mammals Their small size fertility and great adaptability have contributed to this success Today they are abundant on every continent on oceanic islands and are distributed from the equator to the polar regions *See* EUTHERIA RODENTIA FOSSILS [D D D]

Rodentia fossils

Rodents such as squirrels gophers beavers mice hamsters voles jerboas porcupines chinchillas and many other groups are distinguished by their gnawing incisors Rodentia is the largest mammalian order for it includes 42 families and more than 900 genera Rodents range in size from small living mice about 3½ in long to the extinct giant dinomyid *Eumegamys* from the Pliocene in South America which is thought to have been as large as a black rhinoceros On the whole the fossil record of the order is inadequate phylogenies have been outlined for only a few families

Fossil forms like those living have enamel confined to the anterior and anterolabial corner of the incisors incisive foramina of medium or small size strong masseter and temporal muscles with the masseter divided into three major parts The lateral surface of the maxillary is without fenestrations the cheek teeth are low or high crowned with remarkable morphological diversity The basic dental formula is

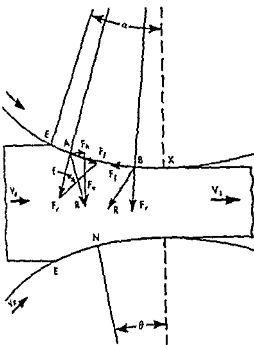


Fig 2 Forces acting between rolls and material

tion and f is friction angle Therefore, stock cannot be drawn into the rolls when contact angle γ exceeds friction angle f The friction angle can be increased by roughening or grooving the rolls, which will increase the effective value of μ

Roll velocity V_R must exceed the velocity V_0 of the material entering the rolls Also, velocity V_1 of the material leaving the rolls must exceed the roll velocity Thus there is a line across the rolls called the no-slip line, where the material velocity and roll velocity are the same Frictional forces on either side of the no-slip line are in opposite directions as at A and at B The position of the no-slip line (point N on diagram) depends upon the balance of horizontal forces F_h No-slip angle θ is related to the friction angle, the no-slip line moves toward the entrance to the rolls when the friction is increased and toward the exit when the friction is decreased or when the degree of reduction is increased. If the no-slip line reaches exit point Y the metal will not be drawn into the rolls

Hot rolling The rolling of a metal at temperatures above its recrystallization temperature is hot rolling. As no strain hardening results greater reductions in thickness may be obtained to economic advantage For most engineering metals hot rolling is carried out at elevated temperatures where oxidation of the surface occurs readily This leads to scale formation decarburization in the case of rolling steel and rather limited dimensional control A fiber structure also results from the alignment of nonmetallic inclusions in the rolling direction

The first step in hot rolling is to roll a heated ingot into a bloom which is a semifinished product of smaller size than the ingot and which will be further processed into a billet or other shape Most blooming mills are two high reversing type mills Grooves are cut in the rolls to accommodate the ingot and the various reductions in cross sections desired, several grooves of different dimensions being in each roll By reversing the direction of the rolls and turning the metal on its side between passes, the piece may be pulled through the rolls perhaps twenty times before its dimensions are reduced (from, say, 18 in × 21 in to 4 in × 6 in

The reduction of a bloom to a billet and then to a bar is accomplished in a continuous mill To hot-work the stock on all sides, it is turned 90° between every pair of rolls by the use of guides which twist the stock Or the bloom could be rolled into a slab and then into strip or sheet, depending upon the dimensions of the final product The maximum reduction per pass is normally limited to about 50%, depending upon material thickness and roll diameter However, a unique and unusual rolling mill (Sendzimir mill) has been developed which can hot roll 2 in stock down to 1/10 in in one pass This mill consists of a series of small working rolls (about 2 in in diameter) mounted inside a cage which in turn surrounds a large backup roll (Fig 3) As the backup rolls rotate at a speed of approximately 300 rpm the cage rotates in the same direction at about 130 rpm This causes the small working rolls to spin in the reverse direction at about 1200 rpm The working rolls are thus turning in a direction against that of the stock and produce a hammering action on the strip

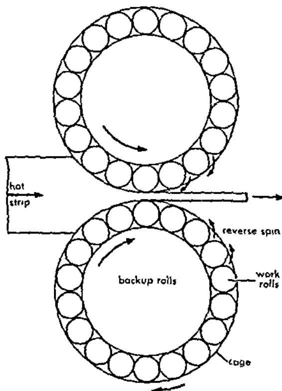


Fig 3 Principles of Sendzimir mill

by ionizing radiation are proportional to the energy absorbed, the roentgen is commonly used in the measurement of ionizing radiations particularly x rays and γ rays, in the study of effects of radia tion on living tissue The roentgen is also used in measurements to determine safe conditions for per sons working in the presence of ionizing radiation The roentgen is defined as the amount of x radia tion or γ radiation such that the associated corpus cular emission per 0 001293 gram of air produces, in air, ions carrying one electrostatic unit of charge of either sign This amount of radiation delivers 83 ergs of energy per gram of air The same radiation will deliver varying amounts of energy to body tis sue, ranging approximately from 90 to 110 ergs per gram, depending on the nature of the tissue

An extension of the roentgen in terms of com parable energy absorption is the rad, defined as the amount of radiation producing an energy absorp tion of 100 ergs per gram The rad is now the offi cially accepted unit replacing the former rep unit which was not always easy to apply in practice

Another unit the rem, defined as the quantity of radiation of any type which when absorbed by man produces an effect equivalent to that produced by one roentgen of 250 kilovolt x rays is not officially accepted because of its vagueness *See* RADIATION BIOLOGY [L F CS]

Roll mill

A series of rolls operating at different speeds Roll mills are used to grind paint or to mill flour In paint grinding, a paste is fed between two low speed rolls running toward each other at different speeds Because the next roll in the mill is turning faster it develops shear in the paste and draws the paste through the mill The film is scraped from the last high speed roll For grinding flour, rolls are oper ated in pairs, rolls in each pair running toward each other at different speeds Grooved rolls crush the grain, smooth rolls mill the flour to the desired fine- ness The term roller mill is applied to a ring roller mill *See* GRINDING MILL [R M H]

Rolling, metal

Deformation of metal by compressive forces ex erted by rolls The most important rolling processes convert ingots into rods strips sheets tubes and structural shapes Rolls may also be used for spe cial applications such as roll forging roll bending and roll forming *See* FORGING, SHEET METAL FORMING

The purpose of rolling is primarily to change the shape of cast ingots into finished products more quickly and more economically than could be done by other means such as forging For some nonfer rous metals bars, tubing, and structural shapes are more economically produced by extrusion processes (*see* EXTRUSION) In addition to the change in shape of the ingot, the rolling operation also re- fines the coarse, as cast structure of the ingot In cold rolling strength is increased by strain harden ing. A smooth finish and close tolerances can also be obtained

Types of mill. The equipment used for rolling depends upon such factors as size and shape of the finished product, amount of reduction to be attained in a single pass through the rolls and type of metal being rolled Grooved rolls are used for rolling the ingot down to intermediate shapes such as blooms and billets, and for structural shapes. Cylindrical rolls are used for producing sheet and strip The simplest type is the two high mill consisting of two equal sized rolls arranged parallel and horizontal one above the other with the rolls rotating in op- posite directions (Fig 1) The use of a three-high mill where the top roll rotates in the same direc tion as the bottom roll eliminates the necessity of reversing the direction of rotation of the rolls if the stock is to be passed back and forth through the same mill Power requirements can be reduced if small working rolls are used but to compensate for the excessive elastic deflections of small diameter rolls larger backing rolls must be used, resulting in some cases in a cluster of as many as five backing rolls for each working roll (called a six high cluster mill) To avoid passing the stock back and forth through the same mill a series of mills arranged close together in a line (continuous mill) may be used, however special care must be taken in the speed adjustment at each mill to compensate for the increased velocity of the strip or rod as it is successively reduced in cross section at each set of rolls Rolling mills may also be equipped with de- vices which guide the stock through the rolls apply a back or front pull on the stock and coil the fin ished product

Theory of rolling As the rolls rotate frictional forces such as F_f (Fig 2) develop between the rolls and the material being rolled Radial forces F_r due to the rolling pressure depends on the friction initial and final thickness of stock roll diameter, and properties of the stock Horizontal component F_h of resultant R between F and F_f is the force tending to draw the material into the rolls The magnitude of this force at the point of entry E determines whether the stock will enter the rolls. At point E

$$F_h = F_f \cos \gamma - F_r \sin \gamma$$

Entry becomes impossible when $F_h = 0$ or when $F_f \cos \gamma = F_r \sin \gamma$ or $F_f/F = \tan \gamma$ But $F_f/F = \mu = \tan f$ where μ is coefficient of fric

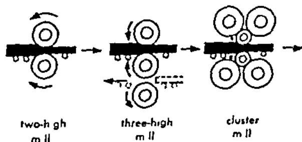


Fig 1 Typical roll arrangements in rolling mills.

ed, 1951 G Sachs and K. R Van Horn, *Practical Metallurgy* 1940 L. R Underwood, *The Rolling of Metals* 1950

Rolling contact

Two bodies are in rolling contact when their relative velocity at the point or along the line of contact is zero. Common applications of rolling contact are the friction gearing of phonograph turntables and speed changers Various instruments and controls utilize such contact An understanding of rolling contact is essential in the study of ball and roller bearings Finally the concepts of rolling contact are helpful in the action of toothed gears

Pure rolling contact may exist between two cylinders—with either internal or external contact—rotating about their centers Two friction disks have external rolling contact when no slipping occurs between them (Fig 1) When any two bodies rotate about fixed centers, Kennedy's theorem (the law of three centers) indicates that the three instant centers for the system lie along a straight line Thus when two bodies in rolling contact are rotating with respect to a third body their point of contact—an instant center—must lie along the line connecting their centers (Figs 1 and 2) In Fig 2 the location of the point of contact P may shift along line $\overline{AB}$ but it must always lie on that line so long as rolling contact exists between bodies 2 and 3

Two curves in driving contact have their angular velocities inversely proportional to the segments into which the line of centers is divided by the common normal to the mating profiles through the point of contact. This fact, along with the law of three centers, indicates that circular arcs rotating about their centers are the only curves that can maintain a constant angular velocity ratio with pure rolling contact.

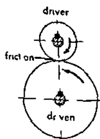


Fig 1 Rolling friction disks

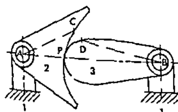


Fig 2 Rolling contact with varying velocity ratio

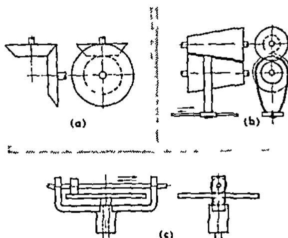


Fig 3 Friction gears. (a) Cones (b) Variable speed (c) Mechanical integrator (International Textbook Co)

To maintain rolling contact the length of the arc of contact along each body during any given time interval must be the same In Fig 2, if bodies 2 and 3 are to maintain rolling contact and if point C is to contact D, the arc distances $\overline{PC}$ and $\overline{PD}$ must be equal Also when A and B are fixed centers, the sum of the radii to the point of contact must be the same for all positions of bodies 2 and 3 Since the lengths of the arcs of contact must be equal on each body continuous rolling contact (each body moving through one or more complete revolutions about its center of rotation) will ordinarily demand that the periphery of each body be the same An *exception to this rule is the case of bodies having* circular cross-sections through the point of contact and perpendicular to their respective axes of rotation (rolling cones) Other geometric shapes such as equal hyperboloids rotating about skewed axes give a combination of rolling and sliding contact Equal ellipses each rotating about one of its foci and having the distance between centers equal to the major axis can have rolling contact

Since pure rolling contact implies equal linear velocity for the point of contact on each body, a definite relation exists between the radii to the point of contact and the angular velocities of the bodies Referring to Fig 2, velocity of P on each body is

$$V_{P2} = \overline{AP}\omega_2 = \overline{BP}\omega_3 = V_{P3}$$

or

$$\frac{\omega_2}{\omega_3} = \frac{\overline{BP}}{\overline{AP}} = \frac{R_3}{R_2}$$

Thus the angular velocities of bodies in rolling contact are inversely proportional to the radii to the point of contact

Friction gearing includes mainly rolling cylinders cones and disks Spring loaded bearings may be used to reduce slippage Because of the limited contact surfaces friction gearing is generally limited to transmission of low torques (Fig 3) Friction devices include friction cones analogous to bevel gearing (Fig 3a) Evan's friction cones for

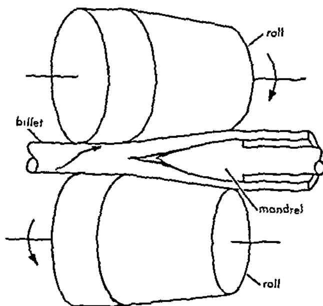


Fig 4 Piercing rolls with axes at an angle to each other spin a solid *billet as they advance it* over a mandrel to produce seamless tubing

Cold rolling. *To maintain* close dimensional tolerance and obtain a smooth surface, the metal must be cold *rolled Theoretically, cold rolling could be* carried out at any temperature below the metal's recrystallization temperature, however the process is *normally accomplished* at room temperature Cold rolling is always accompanied by strain hardening, *which* causes an increase in strength and a *decrease in ductility* A preferred orientation of the crystal lattice also occurs resulting in anisotropic mechanical *properties* (higher strength and lower ductility in a direction transverse to the rolling direction)

Residual stresses generally result from cold rolling and are due to the nonuniformity of plastic deformation across the thickness of the stock The nature of the stresses depends primarily upon the amount of reduction per pass If a light reduction is *taken* the surface layers of the stock are plastically extended in the rolling direction to a greater degree than the core This results in residual compressive stresses at the surface On the other hand if a very heavy reduction is taken in one pass frictional effects at the contact surfaces cause the surface layers to be held back, resulting in greater plastic extension in the core and consequently residual tensile stresses at the surface

The pressure required to cold roll a metal depends upon the strength of the metal, the coefficient of friction between the rolls and the stock and the ratio of contact arc length L to metal thickness t As ratio L/t increases frictional influence becomes proportionately greater in restraining plastic deformation, thereby requiring greater pressure to cause a permanent thickness reduction Thus the apparent yield stress of the metal is raised as the L/t ratio is increased The separating force, which is the average roll pressure times the contact area, increases (1) when the coefficient of friction increases, (2) when the material rolled becomes thinner, (3) as the hardness of the metal rolled increases, and (4) when the roll diameter is increased Eventually a point is reached where the separating force becomes so great that no reduction is possible, the material merely being elastically deformed as it passes through the rolls (the rolls and other machine parts also elastically deforming) To continue rolling, one of the following may be done lubricate the rolls to reduce the friction, anneal the metal to soften it, pack roll (stack sheets and roll together thereby reducing effective L/t ratio), put tension on the strip, or use smaller diameter rolls (to reduce L/t ratio) Excessive lubrication can move the no slip line to the exit [illegible]

by roll piercing, wherein a hot, round billet is fed through two barrel shaped rolls which are rotating in the same direction, but whose axes are at an angle to each other (Fig 4) This causes the billet to spin and advance over a mandrel held between the rolls The rolling action causes the center of the billet to rupture due to secondary transverse tensile stresses, which occur when round stock is compressed in a radial direction Forcing the billet over the mandrel enlarges this center cavity and produces a seamless tube The pierced tube is then sized in subsequent rolling operations

Special rolling processes. Metal powders *may* be rolled into sheets or strips by feeding the powder onto a carrier (such as a strip of paper), hot or cold rolling the powder and carrier together to form a strip and heating the pressed powders to sinter them The rolling and heating cycle are repeated to achieve the required density The carrier may be peeled off or burned off in the first sintering operation Advantages of rolling metal powders include maintaining very fine grain size rolling sheets with no crystallographic orientation (grains are randomly oriented) obtaining purer metal sheets and cladding a metal carrier on one or both sides The cost of equipment for rolling metal powders is relatively small

A process known as Tube in Strip produces a sheet of solid metal in which tubes are produced by inflating existing laminations in the sheet During production of the metal ingot rods of friable material are inserted in the ingot mold During subsequent rolling the rods crush to a fine powder By inserting needles into the ends of the channels that contain the powder and applying air or hydraulic pressure the tubes are produced the shape being controlled by use of profile dies The product is particularly suitable for heat exchangers

A special rolling process for cold forming a long thin walled tube from a short thick walled tube called cold power spinning (or by trade names such as Hydrospin) is in reality a combination of rolling extrusion and spinning *See* METAL FORMING, SHEET METAL FORMING [B L F]

Bibliography J M Camp and C B Francis *The Making, Shaping, and Treatment of Steel,* 6th

nerging from the embryo of a germinating seed is ... mary root It may grow indefi

... root or its branches are known as adve... roots Most frequently these come from stems, but in some plants they may originate from leaves The prop roots of corn (*Zea*), screw pine (*Pandanus*), and mangrove (*Rhizophora*) are examples of adventitious roots Adventitious roots, all of similar size, form the entire root system of most grasses, which thus appear as a mass of long slender structures Such systems are commonly called fibrous roots Taproots and others becoming enlarged in connection with the storage of food are called fleshy roots

Duration of roots. Roots developing from plants which live for one growing season only are said to be annual For example, many weeds and grasses including grains such as wheat, corn, and oat have annual roots (*see* ANNUAL PLANTS)

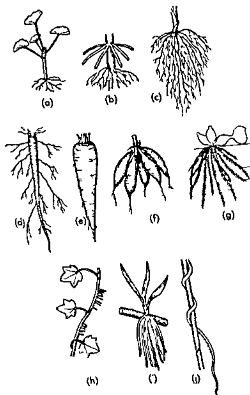


Fig 1 Kinds of roots (a) Adventitious on a cutting or slip (geranium) (b) Prop and other adventitious (corn) (c) Fibrous adventitious (grass) (d) Primary, or taproot (dandelion) e) Fleshy taproot (carrot) (f) Fleshy fascicled adventitious (dahlia) (g) Aquatic adventitious (water hyacinth) (h) Aerial adventitious (English ivy) (i) Aerial adventitious (orchid) (j) Parasitic or haustorial advent t ous (dodder)

Roots persisting for two growing seasons are classified as biennial, as those of beets, carrots, and turnips Such roots accumulate food during the first growing season Although the shoots usually die with the advent of winter, these roots remain alive and the food in them provides the main source of energy and sustenance for the growth and development of new shoots and subsequently the flowers, fruit, and seed during the second and final growing season of the plant *See* BIENNIAL PLANTS

Root systems that continue to grow and function for three or more years are called perennial, as those of trees and shrubs *See* PERENNIAL PLANTS

Many plants have annual shoots, but biennial or perennial roots, as well as roots of plants that are completely annual in temperate regions may persist for more than one year when these plants are grown in tropical areas

Media in which roots grow. Most roots are terrestrial (growing in soil), but some are aquatic (growing in water), and others are aerial (growing in air) The holdfast roots of English ivy are considered aerial roots Also many tropical plants, such as epiphytic orchids and bromeliads growing on the branches of trees, have aerial roots with a specialized surface tissue, the velamen This tissue is presumed to absorb atmospheric moisture or to protect the root from loss of moisture. Parasitic plants, such as dodder (*Cuscuta*), have specialized rootlike haustoria which penetrate the host plant and absorb water and nutrient substances [N A]

Root systems. The extent of the system or root mass varies in relation to inherited growth characteristics and factors such as soil porosity, aeration, and the availability of water In general, root systems are extensive, spreading widely or penetrating deeply, the numerous finer branches and

Table 1 Dimensions of root systems*

Plant	Lateral spread ft	Depth, ft
Little bluestem, *Andropogon scoparius*	1 0	4 0
Blazing star *Liatris punctata*	9 0	16 0
Comanche cactus, *Opuntia camanchica*	9 0	2 8
Wheat *Triticum aestivum*	2 0	5 0
Corn *Zea mays*	8 0	7 0
Sugar beet, *Beta vulgaris*	3 0	6 0

* Based on data from J E Weaver, *Root Development of Field Crops* McGraw Hill, 1926

Table 2. Quantitative characteristics of root systems*

	Crested wheat grass	Winter rye	Coffee
Age	2 years	4 months	3 years
Soil mass	56 cu ft	2 cu ft	270 cu ft
Total root length	315 miles	387 miles	14 6 miles
Number of roots		13 000 000	
Total surface		2 554 sq ft	

* Based on data from various sources as summarized by P J Kramer *Plant and Soil Water Relationships*, McGraw Hill, 1949

transmitting variable speeds by shifting the belt position (Fig. 3b), and the brush wheel and plate (Fig 3c), an arrangement that permits change of velocity ratio as well as sense of rotation of the wheel [R C F]

Bibliography V L. Doughtie and W H James *Elements of Mechanism*, 1954.

Roof construction

In selecting a roof system the important factors to consider are length of span weight, heat insulating value appearance of the undersurface ease of maintenance and cost The roof system consists of joists rafters or purlins the structural deck, and the roof covering

Wood asbestos and asphalt shingles, slate or roofing tile should not be used on roofs with a slope less than 6 in vertical to 12 in horizontal unless special precautions are taken to insure water tightness Pitched roofs are almost always framed in steel or wood The roof deck is supported on purlins or rafters The lower ends of the rafters rest upon the wall plate and their upper ends are supported by the ridge or by other rafters Each end of a purlin rests upon the top chord of a roof truss See TRUSS

A flat roof consists of a wood steel or composition roof deck supported on steel or wood joists The deck is surfaced with a built up roofing formed by cementing together several layers of felt and topping with slag or gravel

Thin shell roofs Thin shell construction is a structural method of enclosing space in which the supporting structure as well as the enclosing structure is in the membrane A shell is a rigid curved membrane whose basic resistance to external forces is through tensile and compressive stresses Bending moments and shears frequently of sufficient magnitude to control the design occur in the region adjacent to the boundary

The three general classes of structural shells are the spherical or dome-shaped cylindrical or barrel shaped (Fig. 1) and hyperbolic paraboloid The



Fig 1 Intersecting barrel shells roof the Administration Building at the St Louis airport Concrete is covered with an insulated plywood deck topped with copper sheeting The underside is covered with acoustical plaster



Fig 2 A folded plate roof of thin shell concrete is cantilevered over the sidewalk at this department store outside Tampa Florida

surface of a spherical or a conoidal dome is described by revolving an arc of a circle about an axis parallel to its radius An elliptical dome is described by revolving an ellipse around its minor axis One of the largest of its type the spherical shell for the University of Illinois arena and auditorium has a 400-ft free-span diameter and a minimum thickness of 3½ in

One of the shapes that adapts well to thin shell construction is the hyperbolic paraboloid whose surface is formed by straight sections Intersecting hyperbolic paraboloids and segments of hyperbolic paraboloids connected together and cantilevered from a common support are just two of the possibilities of this graceful form of design

The corrugated or hipped plate roof and the folded plate roof are forms of construction that are not pure shells because their surfaces are not curved (Fig 2) However with these forms of construction large areas can be spanned with very thin sections

Roof stresses Each member of the roof system must be designed to withstand the maximum stress likely To determine this stress it is necessary to know the magnitude and position of the load The loads to be considered are weight of the roof system weight of snow on the roof wind and in some localities forces due to earthquake Snow load should cover only that portion of the roof necessary to cause maximum stress in the member under consideration The magnitude of the wind load depends upon the slope of the roof There may be pressure or suction on the windward slope and there is always suction on the leeward slope [C N G]

Root (botany)

The root is the absorbing and anchoring organ of vascular plants, it bears neither leaves nor flowers and usually is subterranean The first structure

from the meristematic tip develop into the primary t ssues of the root A terminal *root cap covers* the meristematic region (Fig 2*a*) Although *cell* divi sion can be observed at *some* distance behind the root cap the central cells adjacent to it and their *most recent* derivatives are commonly designated as either the apical meristem or root apex (*see* MERISTEM APICAL)

Apical meristem The apical meristem adds cells to both the root cap and the root proper It is com posed of initials—cells in terminal position—and their immediate derivatives cells in subterminal *position. The arrangement* of the initials varies in different plants Commonly they occur *in two or* more horizontal rows or tiers which have specific developmental relationships with the primary tis sue regions and the root cap In many plants how ever the initials are not arranged in tiers and form *a meristematic region common* to all primary parts of the root. In some lower vascular plants such as most ferns all root tissues arise from a single mer i 1 cell Several studies on roots of

of this cup Most of the apical cells lying w the inverted cup appear to divide rarely during normal growth of the root Studies with radioiso topes support this view Dividing nuclei accumulate certa n isotopes Autoradiographs of sections from root tips of the few species studied have shown little or no radioactive substance in the cells at apex whereas radioactivity was pronounced in the subapical regions and root cap where dividing cells are frequently encountered

Root cap During root growth in the soil cells are lost from the root cap surface as others are added by the apical meristem The outer cells en large as they *mature* and the walls become highly mucilaginous The passage *of the root* tip through the soil is believed to be facilitated by the mucilagi nous substances

Primary meristems Cell divisions are most fre quent *in* the primary tissue regions *of the root* which usually become distinct *immediately behind* the apical meristem In recognition of their meri stematic nature at this root level these regions are often designated as primary meristems These meri stems are the protoderm ground meristem and *procambium Protoderm* represents *the surface* layer which develops into the epidermis procam bium designates the central region which forms the vascular cylinder (stele) and ground meristem constitutes the remaining tissue which gives rise to the root cortex

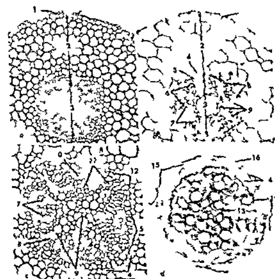


Fig 3 Roots transverse sections showing primary and secondary tissues (a) Soybean (b) Barley (c) Soybean enlarged portion of (a) illustrating early formation of cambium (d) Soybean with abundant secondary tissues 1 epidermis 2 cortex 3 central cylinder stele 4 endodermis 5 pericycle 6 protophloem 7 metaphloem 8 protoxylem 9 metaxylem development of central cells not complete 10 primary phloem fibers 11 cambium 12 divided pericycle cells 13 secondary xylem; 14 secondary phloem; 15 remnants of cortex; 16 periderm

Epidermis The root epidermis *usually* consists of a single layer of elongated thin walled cells in which the cuticle is difficult to demonstrate with certainty (Fig 3*a b*) In many older roots the epi dermis is sloughed off If the layer persists in ma ture roots, more or less conspicuous *cutinization* or other wall *modification* may develop The produc tion *of* root hairs is the most outstanding feature of the root epidermis (*see* EPIDERMIS PLANT)

Root hairs These structures develop as narrow tubular outgrowths of epidermal cells, their walls and contents being continuous with *those* of the main body of the cell They *increase* the absorbing *surface of the root* by contact with a much greater *volume* of soil than would be touched by the root surface without hairs All epidermal cells may be capable of producing root hairs or in some species they may be formed only from special *short cells* called trichoblasts Root hairs are usually confined to a short length *of root* behind the zone of elonga tion Those toward the root tip are youngest and toward the base they are progressively older and longer As a root grows in length new root hairs are initiated toward the tip while the oldest ones usually degenerate The newly developing root hairs make contact with additional soil particles and maintain an extensive absorbing surface

Cortex Thin walled cells among which intercel lular *spaces* develop to varying degrees are com *mon features* of the root cortex Cells may be ar ranged more or less irregularly or they may show radial or *concentric* patterns separate *or in* combi nation Cell arrangement may reflect some aspects of cortical growth for example radial seriation arises as the result of *centripetal* growth involving *successive* pericl nal (parallel with the circumfer ence) divisions of the innermost cells If these *divi* sions are synchronized rings of cells or a concen

their root hairs being in contact with a large volume of soil Their physical attributes are evaluated in different ways For example, as much as 30 feet of root with an external surface up to 65 square inches have been found in 1 cubic inch of soil from under grass In many species the root system constitutes the bulk of the plant body Although most of the roots of many trees may be found in the top 4–5 feet of the earth's crust, their lateral spread may extend through the soil considerably beyond the tips of the longest branches of the crown The form of root systems differs with the origin and manner of growth of their members Nearly all are variations of taproots and fibrous roots

Taproot In taproot systems the primary root forms a dominant central axis The main root penetrates rather deeply into the soil and is generally larger than its branches Plants with taproots range from trees, in which the primary root and its main branches are thick and woody to herbs in which the taproot may be rather slender or develop into a fleshy food storage organ as in a carrot plant

Fibrous roots Several to many main roots of equal dominance typify fibrous root systems Most commonly these main roots arise adventitiously from the stem as in grasses, but sometimes fibrous roots are composed of branches of a primary root that ceased to be dominant In some species root systems organized like fibrous roots are composed of thick and fleshy units, for example, those of dahlia

Root cuttings. Roots of many plants normally form buds from which shoots develop, or they may be induced to form buds by injury or severe pruning Cuttings, or short lengths, of roots may be planted to propagate such plants (*see* REPRODUCTION, PLANT)

Primary tissues of roots. Roots are characterized by a pattern of apical growth similar to that of stems *See* STEM (BOTANY) Cells formed by meristematic activity (cell division) in the root tip are added to the body of the root The addition of the cells and their subsequent elongation result in growth in length of the root Cells derived

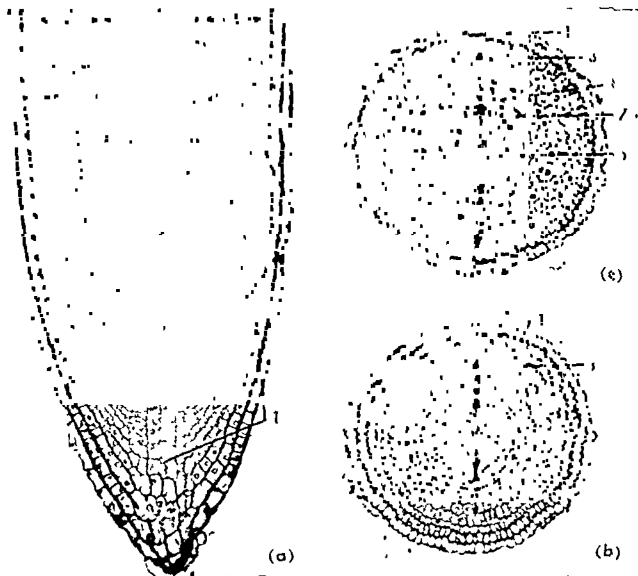


Fig 2 Root tip organization of a tomato plant (a) Median longitudinal section (b c) Transverse sections of same root at 160 and 480 microns behind apex respectively 1, root cap 2, apical meristem 3 epidermis (protoderm) 4 cortex (ground meristem) 5, central cylinder (procambium) 6 protophloem 7, pericycle 8, endodermis.

ing completion When viewed in transverse section the cambium first appears as short disconnected arcs of periclinally dividing cells internal to the phloem strands As vascular cells are formed by the periclinal divisions in the arcs of cambium peri cycle cells opposite the xylem poles also divide periclinally Inner cells formed by divisions in the

ary phloem are deposited inside and ... cambium respectively (Fig 3d) As cambial ac tivity progresses the outer tissues are subjected to ten ion and pressure Primary phloem is crushed and individual cells degenerate except for those which in some plants differentiate into primary

in circumference to diameter increase and ... continues to function throughout the life of the root.

Al ho igh monocotyledons generally lack cambial

duces a secondary body of vascular bundles em bedded in parenchyma tissue

Secondary vascular tissues The cambium and econdary tissues derived from it are similar in gen eral organization to those of the stem but certain quantitative differences between the root and stem tissues and their constituent cells are present One of the most common characteristics of roots is a relatively large volume of parenchyma a feature considered to be a result of high degree of speciali zation for storage The abundant parenchyma re sults either from normal cambial activity or from anomalous growth involving irregular cell proli eration in the secondary xylem or formation of ad ditional layers of cambium outside the normal

Periderm Periderm a secondary protective tis ... characteristic of woody roots but it may not

most commonly has its origin in the pericycle Its characteristic component is the cork tissue *See* BARK PERIDERM

After the vascular cambium becomes complete outside the xylem poles periclinal divisions in the pericycle continue around the vascular cylinder and the outer cells thus formed function as a phel logen It begins to form periderm as the vascular core enlarges and stretches the cortex which is ultimately split and shed If the original periderm is ruptured by diameter growth new layers are formed from new phellogen which differentiates in parenchyma of the secondary phloem

Periderm formation may also follow a different course In some dicotyledons it first arises from subepidermal cells but subsequently another is formed from a proliferated pericycle In others it originates only in the outer cortex Many roots with little or no secondary thickening lack a periderm the outer protective tissue being the persisting cutinized epidermis or an exodermis of subepider mal origin The exodermis comprises layers of cells with suberized or lignified wall thickenings

Root and shoot connection The more or less cylindrical vascular system of the root is morpho logically distinct from that of the shoot which is composed of interconnected vascular bundles In the hypocotyl (seedling axis below the cotyledons) the vascular tissues are arranged in a pattern that is intermediate or transitional between those of the two systems root and stem Because of its inter mediate structure the hypocotyl is often designated as the transition region In this region the rather solid vascular core of the root with its radial tissue arrangement and exarch xylem is merged with separate collateral vascular bundles having en darch xylem These bundles are the leaf traces of the cotyledons the first foliar organs of the shoot Structural features of the transition region are re lated to spatial adjustments which provide tissue continuity between the root and shoot systems of contrasting arrangements of tissues and patterns of differentiation

The organization of the transition region varies greatly in different species Essential features of a

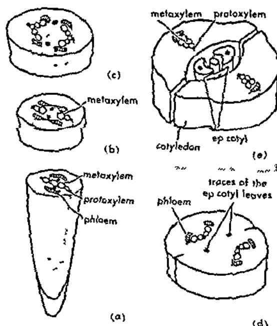


Fig. 4 Diagrams of vascular trans tion in dicotyledon ous seedl ng (beet *Beta vulgaris*) from root tip up ward to level of cotyledons (a–e) Root (a) is diarch its vascular t ssues are cont nuous w th those of the cotyledons (e) through the trans t on reg on (b c d) in the hypocotyl (K Esau Plant Anatomy Wiley 1953)

tric pattern may also be apparent Small inter cellular spaces schizogenous (splitting of adjacent cell walls) in origin develop near the root apex particularly in the inner cortex Larger lysigenous (dissolution of cells) cavities may also develop in the older root One or more layers of the outermost cortical cells may develop suberized or lignified walls which serve protective functions Those of the innermost layer usually become specialized as an endodermis (*see* CORTEX PLANT)

Endodermis Cells of the endodermis exhibit wall modifications which involve thickening and deposi tion of suberin or lignin After the centripetal growth of the cortex is completed suberin and lignin are deposited in the form of a band the Cas parian strip on and within the transverse and radial walls of the endodermal cells In many roots the endodermal cells undergo no further changes in others an extensive thickening of the radial and inner tangential walls follows (*see* ENDODERMIS)

Pericycle The outer portion of the central cylin der or stele is designated as the pericycle which may be one or more cell layers in thickness Al though generally continuous it may be interrupted by xylem elements as in roots of some grasses (Fig 3*c*) Cell divisions in the pericycle initiate the growth of lateral roots and contribute to the formation of the vascular cambium and periderm in roots with secondary growth (*see* PERICYCLE)

Vascular system The vascular system consisting of primary xylem (water conducting) and phloem (food conducting) tissues constitutes the bulk of the central cylinder in many roots Phloem occurs in strands near the periphery xylem alternates with phloem either as strands or as ridges of a central xylem mass This distribution that is radial arrangement of xylem and phloem is characteristic of roots Whereas a central pith occurs in the roots of some dicotyledons it is common among mono-cotyledons although in many of the latter the cen tral core consists of lignified parenchyma including one or more strands of water conducting cells

The outer xylem in the ridges or strands is the first to mature and is designated as protoxylem the inner matures after the protoxylem and is called the metaxylem Thus the maturation of the primary xylem occurs in a centripetal direction and the xylem having such a pattern of maturation is termed exarch Protophloem is likewise external to metaphloem As seen in cross sections the points of origin of the protophloem and protoxylem may be called the phloem and xylem poles respectively The number of poles varies among roots of different plants different roots of the same plant or in dif ferent levels of the same root Roots are described as monarch diarch and triarch on the basis of whether there are one two or three poles those with many poles are designated as polyarch Dicot yledon roots generally show fewer poles than those of the monocotyledons (*see* PHLOEM XYLEM)

Lateral roots Root branches or lateral roots are generally produced at some distance behind the apex and they originate endogenously or deep within the tissues arising near either phloem or xylem poles or between them A lateral root is com monly initiated by localized cell divisions in the pericycle however in lower vascular plants the initial divisions may involve endodermal cells Con tinued divisions produce a primordium in which the root cap apical meristem and primary tissue re-gions are soon organized The young root grows through the cortex and by the time it emerges on the surface vascular connections with the main axis are established

Elongation Some elongation characteristic of primary growth of the root results from the in crease in number of cells in the meristematic tip but most of it is produced by the elongation of cells behind the meristematic region The location and extent of the zone of elongation varies in part with root size and stage of growth For example in rela tively thin rapidly growing roots of timothy (*Phleum*) the zone of cell elongation occurs 450–1400 microns behind the root apex whereas in larger primary roots of corn (*Zea*) the zone lies 2000–9000 microns behind the root tip

Differentiation The primary tissues of the root begin to differentiate during the elongation but complete their differentiation or become mature after the elongation ceases Differentiation consists in a series of changes through which the cells be-come more or less specialized with regard to the functions they carry out in the mature state The differentiation follows characteristic patterns in both longitudinal and transverse directions Thus for example the protophloem matures closer to the apex than the protoxylem and both differentiate in the direction from the base of the root toward the apex that is acropetally The metaphloem and metaxylem follow in the same direction As seen in cross sections *both phloem and xylem mature in a* centripetal direction that is from the periphery of the vascular cylinder toward its center The phe nomena of growth that is cell division and cell en largement overlap with the earlier phenomena of differentiation so that the sequence of maturation of a root is highly complex

Secondary tissues of roots The formation of secondary xylem and secondary phloem by the vascular cambium is the chief characteristic of sec ondary organization in roots (*see* MERISTEM LAT ERAL) Other structural modifications invariably accompany cambial activity and thus are also fea tures of the secondary state

The extent of secondary tissue formation in roots generally parallels that in stems Trees and shrubs with woody stems are known to develop woody roots Herbaceous plants with a smaller volume of secondary tissues in the stem also show less cam bial activity in the roots However roots of plants that have little or no secondary vascular tissues in the stems have not been investigated widely to determine the presence or absence of a cambium

Development of vascular cambium Formation of the vascular cambium in the dicotyledons begins at root levels where primary development is approach

minimizing the root mean square of the errors accomplishes the same purpose *See* CURVE FITTING, STATISTICS [H R C]

Rope

A long flexible structure consisting of many strands of wire, plastic, or vegetable fiber such as manila. Rope is classified as a flexible connector and is used generally for hoisting, conveying, or transporting loads, transmitting motion, and occasionally for transmitting power. For flexibility, and to reduce stresses as the rope bends over the sheave (pulley), a rope is made of many small strands rather than few large ones.

Wire rope. In making wire rope, the wires are preformed and wound into strands which are then twisted together to form the rope. A hemp center is usually used to hold lubricant for reducing friction and wear as the wires rub over each other when the rope is bent.

Figure 1 shows a 6 × 7 construction (6 strands each with 7 wires). The hemp center is shown in black. Usual constructions are 6 × 7 coarse laid, 6 × 19 standard hoisting, and 8 × 19 and 6 × 37 extra flexible hoisting for elevator service. To keep bending stresses low, sheaves of large diameter must be used. Factors of safety based on ultimate strength of the rope vary from 3 to 8 depending on the application. Some manufacturers' data allow for bending stresses in normal size sheaves; otherwise these stresses must be allowed for separately.

There are two common ways of winding wire rope indicated by lay of the rope. Standard rope is called regular lay, the wires in the strand are laid to the left and the strands to the right. The wires are at right angles to the motion of the rope as it slides into the sheave groove so that the wire may wear excessively, particularly if the sheaves are not well aligned. In lang lay, wires in the strand and the strands themselves are laid in the same direction. This type rope wears better, is more flexible, and lasts longer. Flat rope is made of right and left strands placed alternately beside each other and sewed together with soft steel wire. This rope can be used on a reel slightly wider than the rope; it is used to save space.

For aircraft service 6 × 7, 7 × 7, and 7 × 19 cord and 19 wire strand have been developed. They are made of tinned or galvanized carbon steel or *stainless steel*.

Materials for commercial wire ropes in order of strength and cost are iron, mild plow steel, plow steel, and improved plow steel. Galvanized wire rope may be used for standing service as for guy wires but should not be used in hoisting because



Fig. 1 Wire rope construction.

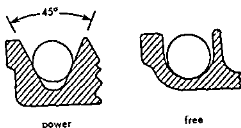


Fig. 2 Groove profiles for rope.

as they bear against the grooves in power transmission (Fig. 2). Internal corrosion due to water or other corrosive substances has caused failure. Profiles of grooves on the sheaves are semicircular at the bottom when the rope runs free for hoisting and transportation (Fig. 2). The rope should not be pinched at the bottom of the groove. For power transmission with wire ropes the contacting part of the semicircular groove should be lined with rubber or similar resilient material.

Wire ropes may be attached to a hoisting drum by using a rope socket at the end of the rope; the socket is slipped into a pocket cast in the drum at the end of the last groove. The socket is held in place by a screwed plug, which can be easily removed to detach the rope. To attach the rope to its load or to a block and tackle anchor, various sockets, clips, and thimbles are available.

Manila rope. Manila fiber is twisted into yarn, the yarn is twisted into strands, and the strands are laid up into rope. Manila rope is used mainly for light service hoisting but sometimes for long distance transmission of power. Being flexible this rope accommodates itself to small size pulleys and absorbs starting and shock loads, but the fiber at the center of the strand powders and pulverizes although the rope may appear sound at the surface. This deterioration, from the rubbing of strands on each other as the rope flexes, can be retarded by such lubricants as acid-free paraffin or graphite worked into the strands during laying or twisting of the strands to form the rope.

For power transmission and for hoisting with *manila rope*, see the groove profiles shown in Fig. 2. The 45° groove angle used for power transmission provides sufficient friction to prevent slippage between the rope and cast iron surface of the groove. [P H B]

Bibliography. W. Kent, Design and Production, *Mechanical Engineers' Handbook*, vol. 1, 1950.

Rope drive

A means to transmit power when the distances are so great as to make the use of belts impractical. Rope drives use ropes running in grooved pulleys or sheaves with the contact between the rope and sheave similar to that between a V belt and sheave. In areas protected from weather, the ropes may be of hemp or similar fibrous material. For long dis

pattern common among dicotyledons are shown in Fig 4 In this example the diarch xylem plate of the root is continuous with two leaf traces of the cotyledons, each obviously double in structure At successively higher levels the relative position of protoxylem and metaxylem changes Each phloem strand of the root is connected with each of the two cotyledonary traces Initial union with the vascular system of the epicotyl (shoot above the cotyledons) is established indirectly with the root, the traces of the first leaves remaining distinct throughout the hypocotyl

Other types of transition occur in plants with more than one cotyledonary trace or with three or more poles of xylem and phloem in the root Differences in the timing of epicotyl development and in the pattern of germination (epigean or hypogean cotyledons) also effect variations in the transition region In plants that develop a cambium, secondary tissues are continuous between the root and stem

The general organization of the transition region in gymnosperms is similar to that of dicotyledons, any variations being related to the usually greater number of cotyledons in gymnosperms and to the number of traces to each cotyledon In monocotyledons the vascular system of the root is commonly connected with those of the single cotyledon and the epicotyl, transitional characteristics may occur in both connections or only in that with the cotyledon A particularly complex transition characterizes the grass family, Gramineae Vascular strands of the root join in a nodal plate at about the level of the scutellum or cotyledon Also joined with the plate are vascular strands of the scutellum coleoptile (first leaflike structure above the cotyledon), and first leaves In this complex pattern the transition between vascular systems of root and shoot is relatively abrupt and the root system is connected with more than one leaf above the scutellum See SEED (BOTANY) [C HE]

Bibliography See PLANT ANATOMY

Root (mathematics)

If a function $f(x)$ has the value 0 for $x = a$, a is a root of the equation $f(x) = 0$ The fundamental theorem of algebra states that any algebraic equation of the form $a_0x^n + a_1x^{n-1} + \cdots + a_{n-1}x + a_n = 0$ where the a_ks are real numbers has at least one root From this it follows readily that such an equation has roots, real or complex, in number equal to the index (here n) of the highest power of x

Furthermore, if $a + ib$ (where $i = \sqrt{-1}$) is a complex root of the given equation so is $a - ib$, the conjugate of $a + ib$ Equations of degrees up to 4 may be solved algebraically This statement means that the roots may be expressed as functions of the coefficients, the functions involving the elementary arithmetical processes of addition, multiplication, raising a number to a power or extracting the root of a certain order of a given number It has been proved by H Abel and by E. Galois that it is not possible to solve algebraically the general algebraic equation of degree higher than four However, it is possible to determine the real roots of an algebraic equation to any desired degree of approximation

The term zero is sometimes used in lieu of root when dealing with functions which are defined as infinite power series Thus, one talks about the zeros of $\sin x$, $\cos x$, and $J_0(x)$ (the Bessel function of first kind and zeroth order) Each of these three functions may be expressed as an infinite power series

Numbers are called transcendentals if they can not be the roots of any algebraic equation with integral or rational coefficients The most important transcendental numbers are π, the ratio between the circumference of a circle and its diameter, and e, the base of the system of natural logarithms See CALCULUS DIFFERENTIAL AND INTEGRAL, EQUATIONS, THEORY OF, NUMERICAL ANALYSIS [A N L.]

Root-mean-square

The square root of the arithmetic mean of the squares of a set of numbers is called their root mean square If the numbers are $x_1, x_2, x_3, \ldots, x_n$, the root mean square is equal to

$$\sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \cdots + x_n^2}{n}}$$

It is valuable as an average of the magnitudes of quantities, and it is not affected by the signs of the quantities

Among applications of root mean square the most important is the standard deviation from the arithmetic mean If

$$\bar{x} = \text{arithmetic mean} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

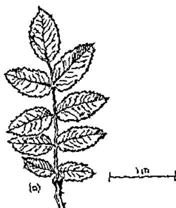
then the standard deviation $= s =$

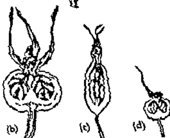
$$\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

Thus standard deviation from the mean is the root mean square of the deviations from the mean It has a great advantage over other measures of deviation because its computation can be greatly facilitated by the following transformation If a set of values $x_1, x_2, x_3, \ldots, x_n$ are each reduced by the same number k and then the results are divided by the same number p, the root mean square of the deviations of the resulting quantities from their mean is equal to $1/p$ times the root mean square of the deviations of the xs from their mean Thus if $x_i = pu_i + k$,

standard deviation of the xs $= p \times$ (standard deviation of the us)

The most common method of minimizing errors in the use of a formula to fit empirical observations is that of determining the coefficients in the formula so as to make the mean of the squares of the errors least This is the principle of least squares Its relation to root mean square is obvious since



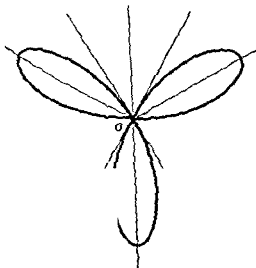


(a) A 9 fol olate rose leaf (b) Form of rose hip Rosa rugosa (c) R pendul na (d) R carol na (L H Bailey The Standard Cycloped a of Hort culture vol 3 Mac m llan 1937)

economic importance The fruits of a few species are used in preserving and the flowers of *Rosa alba* and *Rosa damascena* are used in perfumes especially attar of roses and rose water See ROSALES [P D S]

Rose curve

A type of plane curve that consists of loops (*leaves petals*) emanating from a common point and that has a roselike appearance Taking the common point O as the pole of a polar coordinate



system these curves have equations of the form $\rho = a \sin n\theta$, $a > 0$, and n a positive integer (also $\rho = a \cos n\theta$, with a different choice of the initial line of the coordinate system) The curve is a circle of diameter a for $n = 1$ It has n or $2n$ leaves according as n is an odd or even integer, respectively The area enclosed by one leaf is $\pi a^2/4n$ and so the entire area of the three leaved rose $\rho = a \sin 3\theta$ is $\pi a^2/4$ and that of the four leaved rose $\rho = a \sin 2\theta$ is $\pi a^2/2$ The lemniscate is sometimes called a two-leaved rose though its equation $\rho^2 = a^2 \cos 2\theta$ is not of the form given above. See LEMNISCATE (OF BERNOULLI)

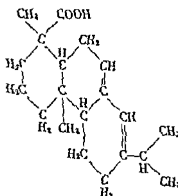
[L M BL]

Rosin

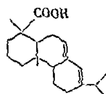
An amorphous, brittle resin obtained mainly from pine stumps by solvent extraction Refining gives a substance ranging in color from a dark brown to a pale lemon yellow

It is insolubl
ganic solvents
and about 10%
ily in air because of its unsaturated components becoming harder and darker in color Rosin softens at about 80–90°C.

The chief resin acids are of the abietic type



or



Abietic acid

The hydrogenated esterified and neutralized products of rosin find application in paper sizing inexpensive lubricants, insulation linoleum adhesives soldering fluxes binders soaps and finishes See TALL OIL TURPENTINE

[E L S]

Rosin oil

A number of commercially useful products are made from *rosin* Some of these may be considered as oils because they are viscous liquids With one exception these so-called oils are esters of abietic acid which is the major constituent of the isomeric acids that make up over 90% of rosin

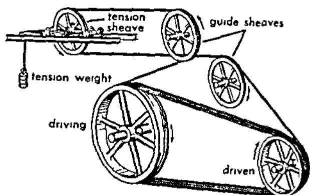
tance drives or installations exposed to the weather, for example, ski tows, wire ropes are sometimes used Rope drives have advantages over belts they can transmit more power over greater distances, they generally require less pulley face width, and shaft alignment is not so critical

Multiple drive. Two systems of rope drives are used In the multiple rope or English system, a number of separate rope loops run side by side in parallel grooves in each sheave In this respect the English system is similar to a multiple V belt drive The number of ropes used depends on the power being transmitted In this drive, contact between the rope and the sheave is maintained by the weight of the rope and the wedging action of the rope in the groove of the sheaves, as a consequence the English system is not suitable for vertical drives A distinct advantage of the English system is that, if one rope fails, the drive is maintained by the remaining loops It is, however, difficult to distribute the load evenly to all loops in the drive For this reason, the wear will not be uniform between loops Since the English system also does not make allowance for the stretch that occurs with use, frequent resplicing is required

Continuous drive. The continuous rope or American system uses one continuous long rope (see figure) The number of turns of rope wound around the driving and driven sheaves depends on the power to be transmitted The rope leaving the last groove of the driven sheave is returned to the first groove of the driving sheave over one or more guide or idler sheaves one of which, the tension sheave is mounted on an adjustable, weighted carriage The weighted carriage maintains the constant tension necessary to secure good contact of the rope and the grooves in the sheaves This system has the distinct disadvantage that, if the rope breaks the entire drive fails The American system however can be adapted to vertical and quarter turn drives whereas the English system is limited to parallel shafts The American system is also more suitable for drives exposed to the weather

Wire rope. When the distances are too great for manila rope as, for example in cable and inclined railways, wire rope is used Two systems with wire rope may be used The continuous or endless rope is used in cable ways The second type single loop is similar to belt drives In some applications the rope does not continually run over both sheaves but instead may be fastened to one of them An example is the rope drive for an elevator or hoist With this arrangement the member to which the rope is fastened is called a drum rather than a sheave or a pulley The effective radius of a drum is increased as each layer of rope is wound onto it

Kinematically, a rope drive is the same as a belt drive, the ideal speed ratio may be calculated in the same manner as for a belt drive The efficiency of rope drives decreases with increased rope speed Typical efficiencies of English system drives are the order of 90% at a rope speed of 2500 feet per minute (fpm) and 85% at a speed of 5000 fpm



American system rope drive (*From V L Doughtie and W H James, Elements of Mechanism, Wiley, 1954*)

The American system is somewhat more effective with efficiencies about 5% higher at each of the previously mentioned speeds *See* BELT DRIVE

While rope drives are the preferred mechanical drive for transmission of power over considerable distance, the use of such drives is becoming limited Electrical energy can be transmitted more conveniently and the growing practice of a separate motor for each machine makes rope drives obsolescent for mill and shop drives The use of rope drives in cableways particularly for material handling continues to be of importance *See* PULLEY, V BELT [R C F]

Bibliography T Baumeister (ed) *Marks' Mechanical Engineers' Handbook*, 6th ed , 1958, V L Doughtie and W H James, *Elements of Mechanism*, 1954

Rosales

A large and greatly diversified order of the plant subclass Dicotyledoneae which includes 11 families having 897 genera and more than 19 600 species The group is of great agricultural horticultural and floricultural value The legume family (Leguminosae) with 550 genera and about 13 000 species is the largest Second in size is the rose family (Rosaceae) with 115 genera and 3200 species This order has representatives in a wide range of habitats in almost all areas of the plant world *See* DERRIS LEGUME FORAGES, LICORICE ROSE *see also* DICOTYLEDONEAE EMBRYOPHYTA PLANT KINGDOM [P D S]

Rose

A member of the genus *Rosa* of the rose family (Rosaceae) These plants are widely distributed in the temperate regions and on tropical mountains The number of floricultural varieties and hybrids is over 3000 and new ones are being added each year Roses are erect climbing or trailing shrubs generally prickly stemmed and bear alternate odd pinnate (rarely simple) single leaves They usually have large showy variously colored flowers Probably no other flower has had such an important place in the garden and in literature In the United States alone, the value of the blossoms grown for sale is estimated to be more than $6 000 000 annually Except for ornamental purposes roses are of little

cable runs from the draw works over a crown block at the top of the derrick and down to the traveling block. The mud is pumped through a hose attached to the swivel An opening in the center of the swivel permits the mud to pass down through the attached drill string

When the bit has penetrated the distance of a pipe section drilling is stopped, the string is pulled up to expose the top joint, the kelly is dis connected a new section added the kelly attached, the string lowered and drilling resumed This proc ess continues until the bit becomes worn out, at which time the entire drill string must be pulled Pipe is usually disconnected in thribbles or 90 ft sections of pipe, and stacked in the derrick The height of the derrick determines whether doubles thribbles, or fourbles can be stacked The process continues until the bit reaches the surface A new bit is attached and the drilling string reassembled and lowered into the hole Such round trips may take up to two thirds of total rig operating time depending upon depth of the hole In hoisting or lowering the drill string the swivel is disengaged from the hook Elevators, or clamps which grip the pipe securely are attached The elevators are also used when the hole is lined with steel casing In lowering drill pipe or casing each new section of pipe is lifted from the derrick floor and suspended on the elevator until it is screwed to the preceding joint just above the hole opening the entire column is then lowered into the hole While new sections of pipe or casing are being attached to the elevators the pipe in the hole is supported in the rotary table by slips or gripping devices

[illegible] can be skid truck or trailer mounted

[illegible]

is set 7 20 ft or more above the ground to provide a basement for control devices such as blowout preventers below the rotary table See CABLE TOOL DRILL OIL AND GAS WELL DRILLING, TURBODRILL

{[illegible]}

Rotational motion

The motion of a rigid body which takes place in such a way that all of its particles move in cir cles about an axis with a common angular veloc ity also the rotation of a particle about a fixed point in space Rotational motion is illustrated by (1) the fixed speed of rotation of the earth about its axis (2) the varying speed of rotation of the flywheel of a sewing machine, (3) the rotation of a satellite about a planet in which both the speed of rotation and the distance from the center of ro tation may vary (4) the motion of an ion in a cyclotron where the angular speed of rotation re mains constant but the radius of the circular mo tion increases and (5) the motion of a pendulum in which case the particles describe harmonic mo tion along a circular arc

The present discussion of *rotational motion* is limited to circular motion such as is exhibited by the [illegible]

cer[illegible] ICS, EARTH (ORBITAL MOTION), [illegible] TION, PARTICLE ACCELERATOR, PENDULUM

Circular motion is a rotational motion in which each particle of the rotating body moves in a cir cular path about an axis The motion may be uni form, that is, with constant angular velocity, or nonuniform, with changing angular velocity

Uniform circular motion The speed of rotation, or angular velocity, remains constant in uniform circular motion In this case, the angular displace ment θ experienced by the particle or rotating body in a time t is $\theta = \omega t$, where ω is the con stant angular velocity

Nonuniform circular motion. A special case of circular motion occurs when the rotating body moves with constant angular acceleration If a body is moving in a circle with an angular acceleration of α radians/sec², and at a certain instant, it has an angular velocity ω_0, then at a time t sec later, the angular velocity may be expressed as $\omega = \omega_0 + \alpha t$, and the angular displacement as $\theta = \omega_0 t + \frac{1}{2}\alpha t^2$ See ACCELERATION, VELOCITY

Banking of curves When a car travels around a curve on a highway, the path is a circular arc of radius R, where R is the radius of curvature of the roadway In order to have the car move in this cir cular arc a horizontal external force must be ap plied to give the car an acceleration perpendicular to its path that is, toward the center of rotation This force must equal Mv^2/R, where M is the mass of the car, and v its speed This centripetal force is supplied by the friction between the tires and the road (see CENTRIPETAL FORCE) If the force of fric tion is not great enough to produce this accelera tion the inertia of the car will tend to make it con tinue with its speed in a straight line tangent to the road rather than around the curve and this will cause the car to slide off the road

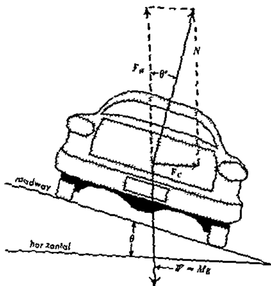


Fig 1 Banking of a curve

Methyl abietate and methyl dihydroabietate are used primarily in coatings, adhesives, transparent izing compounds, plastic compositions, and in miscellaneous applications, such as printing inks and polishes

The glyceryl, ethylene glycol, diethylene glycol, and triethylene glycol esters of hydrogenated rosin are characterized by resistance to oxidation and discoloration They find applications as plasticizers, textile sizings, and as components in lacquer and varnishes, water paints, and pressure adhesives

Hydroabietyl alcohol is a primary, monohydric alcohol It is a resinous plasticizer, and a pigment grinding medium, it is used in plastics, lacquers and other surface-coating preparations The chemical properties of the alcohol lead to its use as an intermediate for further chemical processing and to the preparation of modified resins *See* ROSIN, TALL OIL [E L S]

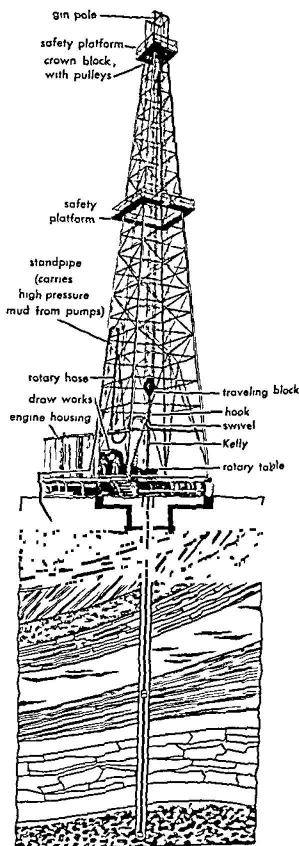
Rotary tool drill

A bit and shaft used for drilling wells A turntable on the derrick floor rotates a string of hollow steel drill pipe at the bottom of which is a steel bit The bit grinds the rock A drilling fluid is pumped down through the drill pipe, the fluid flushes out the rock cuttings and returns up the space between drill string and hole side

The drilling fluid may be air, water, or, most commonly, mud (a mixture of various clays and chemicals, each having a special function) The mud cools and lubricates the bit, removes cuttings from the hole, and cakes the wall of the hole to prevent caving before steel casing is set The hydrostatic pressure exerted by the column of mud in the hole prevents blowouts which may result when the bit penetrates a high pressure oil or gas zone When the mud reaches the surface, it passes over a vibrating screen to filter out large cuttings The mud then passes on to a settling tank where smaller particles settle out The cuttings are examined to determine the type of formation being drilled and for possibilities of oil or gas production The mud mixture is sucked up from the pit and recirculated by a high pressure pump The viscosity weight and filtration properties of the mud are altered as drilling proceeds by changing the proportion of its constituents

Power is transmitted from an engine to a draw works—a winch which drives the rotary table on the derrick floor and also applies power for hoisting or lowering the drill string as shown The string of drill pipe is topped at the surface by a square-sided length of heavy pipe called the kelly The square shape permits the rotary table to grip and rotate the kelly, and hence the entire drill string and yet have sufficient freedom so that it can slip vertically through the table as drilling goes deeper Rotation speeds range from 40 rpm to 500 rpm or more, depending primarily upon the character of the formation being drilled The drill string usually consists of 30-ft lengths of drill pipe coupled together On the lower end are heavier walled lengths of pipe, called drill collars, which help regulate weight on the bit

The drill string is attached to a swivel suspended from a hook which is connected to a traveling block, or pulley, encased in a frame The drilling



Bit is rotated at bottom of well by connected sections of pipe driven from a rotary table atop well

5I

ten quartz, or fused quartz, is ... une is a typical material of the second class It gives a rotation of $-37°$ in a 10-cm length for the

s

1

in most cases for the majority of the ... rotation and should not strictly be termed rotatory dispersion It depends on the fact that *optical activity* is actually circular birefringence *In other* words, a substance which is optically active transmits right circularly polarized light with a different velocity from left circularly polarized light

Any type of polarized light can be broken down into right and left components (see POLARIZED LIGHT) Let these components be R and L The lengths of the rotating light vectors will then be $R/\sqrt{2}$ and $L/\sqrt{2}$ At $t = 0$ the R vector may be at an angle ψ_r with the X axis and the L vector at an angle ψ_l Since the vectors are rotating at the same velocity they will coincide at an angle β which bisects the difference

$$\beta = \frac{\psi_r + \psi_l}{2} \quad (1)$$

If $R = L$ the sum of these two waves will be linearly polarized light vibrating at an angle α to the axes

$$\alpha = \frac{\psi_r - \psi_l}{2} \quad (2)$$

If in passing through a material one of the circularly polarized beams is propagated at a different velocity the relative phase between the beams will change

$$\psi'_r - \psi'_l = \frac{2\pi d}{\lambda}(n_r - n_l) + \psi_r - \psi_l \quad (3)$$

where d is the thickness of the material λ is the wavelength and n and n_l are the indices of refraction for right and left circularly polarized light The polarized light incident at an angle α has according to this equation been rotated an angle

$$\gamma = \frac{\pi d}{\lambda}(n_r - n_l) \quad (4)$$

This shows that the rotation would depend on wavelength even in a material in which n and n_l were constant and which thus had no circular dispersion It is for this reason that the term rotatory dispersion is perhaps ill defined in much of the literature

In addition to this pseudo-dispersion which depends on the material thickness there is a true rotatory dispersion which depends on the variation with wavelength of n and n

From Eq (4) it is possible to compute the circular birefringence for various materials This quant ... order of magnitude of 10^{-8} for many with ...

Bibliography T M Lowry, *Optical Rotatory Power*, 1935

Rotifera

... ing diversity of form and habitat The ... acteristic structure, the corona, is a retractile trochal disk provided with several groups of cilia and located on the head When in motion, the coronal cilia give the illusion of a pair of rapidly rotating wheels especially in bdelloid rotifers The class name Rotifera literally means wheel bearers, and rotifers are often referred to as wheel animalcules.

The classification of rotifers proposed by A Remane is usually followed He divides the class Rotifera into three orders Seisonacea, Bdelloidea and Monogononta

Despite their diversity in form and structure, rotifers are alike in being bilaterally symmetrical, in having several complete organ systems including digestive excretory nervous, and reproductive, and in possessing several structures unique to the class, including the corona and mastax They lack separate respiratory and circulatory systems These animals are similar in regard to their small size, most being between 0 1 and 0 5 mm long, and in regard to their absolute dependence upon water, or at least moisture

The sexes are separate in rotifers Females are more numerous than males and quite different in appearance The material given in this article applies chiefly to females. With but a few exceptions, male rotifers are degenerate, possessing neither mouth nor digestive organs Consequently their life span is a matter of hours to a few days

Morphology The body of a rotifer is usually divided into three parts the head the trunk and the foot. The head carries the corona the mouth and mastax of the digestive system, as well as the central ganglion (brain) of the nervous system Most organs are located in the trunk, this includes the stomach intestine cloaca anus, and gastric glands of the digestive system, the simple excretory system, and the reproductive organs The ex

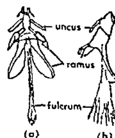


Fig 1 Fulcrate trophus of Seison (a) Dorsal view (b) Lateral view (After de Beauchamp)

In order to reduce the probability of skidding, roadways are customarily banked as illustrated in Fig 1, which shows a car of mass M going away from the reader with a speed v and making a right hand turn along an arc of radius R The road way must exert a vertical force F_W upward, equal and opposite to the weight $W = Mg$ of the car (g is the acceleration of gravity) and a horizontal centripetal force $F_C = Mv^2/R$ to make the car move in a circular arc The net force N of the road on the car is the vector sum of these two forces

From the diagram, it can be seen that the angle θ' which N makes with the vertical is given by

$$\tan\theta' = \frac{F_C}{F_W} = \frac{Mv^2/R}{Mg} = \frac{v^2}{gR}$$

If this angle θ' is equal to the bank angle θ of the road the force N of the road on the car is perpendicular to the roadway, and there will be no tendency to skid This equation shows that the correct bank angle is proportional to the square of the speed and inversely proportional to the radius of the curve For a given curve, there is no correct bank angle for all speeds, thus roadways are banked for the average speed of traffic Bank angle enters into the design of railroads and into the banking of an airplane when it executes a turn

Work and power relations. A rotating body possesses kinetic energy of rotation which may be expressed as $T_{rot} = \frac{1}{2}I\omega^2$, where ω is the magnitude of the angular velocity of the rotating body and I is the moment of inertia, which is a measure of the opposition of the body to angular acceleration The moment of inertia of a body depends on the mass of a body and the distribution of the mass relative to the axis of rotation For example, the moment of inertia of a solid cylinder of mass M and radius R about its axis of symmetry is $\frac{1}{2}MR^2$

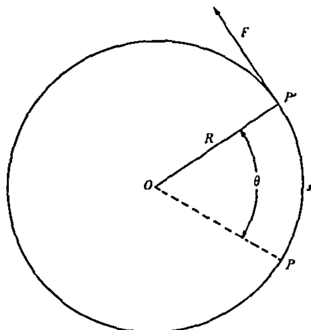


Fig 2 Work in angular motion

In order to impart kinetic energy to a rotating body, work must be done In Fig 2 there is represented a solid cylinder of mass M and radius R, capable of rotation without friction about an axis O By means of a cord wrapped around the cylinder, a constant force F is applied thus imparting angular acceleration to the cylinder If the cylinder is originally at rest and the force F acts through a distance $s = R\theta$ equal to the arc PP', thus rotating the cylinder through the angle θ, the work W done is $W = Fs = FR\theta = L\theta$, where $L = FR$ is called the torque or moment of force The action of this torque L is to produce an angular acceleration α according to the equation

$$L = I\alpha = I\frac{d\omega}{dt} = \frac{d}{dt}(I\omega)$$

where $I\omega$, the product of moment of inertia and angular velocity is called the angular momentum of the rotating body This equation points out that the angular momentum $I\omega$ of a rotating body, and hence its angular velocity ω remains constant unless the rotating body is acted upon by a torque Both L and $I\omega$ may be represented by vectors

It is readily shown that the work done by the torque L acting through an angle θ on a rotating body originally at rest is exactly equal to the kinetic energy of rotation

$$W = L\theta = I\alpha\theta = I\alpha\tfrac{1}{2}(\alpha t^2) = \tfrac{1}{2}I(\alpha t)^2 = \tfrac{1}{2}I\omega^2$$

because for the case at hand, $\theta = \frac{1}{2}(\alpha t^2)$ and $\omega = \alpha t$

Power is defined as rate of doing work, and the power P in rotational motion is

$$P = \frac{dW}{dt} = \frac{d}{dt}(L\theta) = L\frac{d\theta}{dt} = L\omega$$

See ANGULAR MOMENTUM, MOMENT OF INERTIA, POWER, RIGID BODY DYNAMICS TORQUE WORK

[C E H]

Rotatory dispersion

A term used to describe the change in rotation as a function of wavelength experienced by linearly polarized light as it passes through an optically active substance *See* OPTICAL ACTIVITY

Optically active materials Substances that are optically active can be grouped into two classes In the first the substances are crystalline and the optical activity depends on the arrangement of nonoptically active molecular units When these crystals are dissolved or melted the resulting liquid is not optically active In the second class the optical activity is a characteristic of the molecular units themselves Such materials are optically active as liquids or solids A typical substance in the first category is quartz This crystal is optically active and rotates the plane of polarization by an amount which depends on the direction in which the light is propagated with respect to the optic axis (*see* CRYSTAL OPTICS) Along the axis, the rotation is 29 73°/mm for light of wavelength

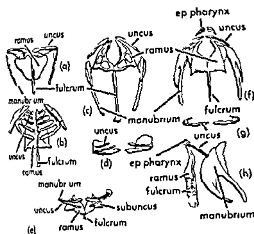


Fig 3 (a) Malleate trophus of a brachionid (schema c) ventral view (b) Upper view (*After de Beauchamp*) (c) Virgate trophus of *Notommata* ventral view (d) Upper view (*After Meyers*) (e) Uncinate trophus of *Stephanoceros* Collothecacea upper view (*After de Beauchamp*) (f) Cardate trophus of *Lindia* ventral view (g) Upper view (h) Lateral view (*After Meyers*)

ing worker on rotifers in the United States found that soft water habitats have a markedly richer fauna, in number of species than hard water lakes and ponds Rotifers occur most commonly in fresh water habitats Only about 12% of the known rotifers have been taken in brackish or marine habitats A few species principally in the genus *Synchaeta* and among the Seisonacea are the only true marine species A few rotifers are epizoic or parasitic A fairly large group of rotifers especially bdelloids live in mosses and lichens even in xerophytic situations Some bdelloids can withstand extreme desiccation lasting even for years Rotifers have been collected in hot springs at temperatures as high as 45°C and include both bdelloid and ploimate species

Most rotifers are short lived having a life span of 1 3 weeks Some of the moss dwelling bdelloids may live 1 5 months or even longer if periods of desiccation intervene See ASCHELMINTHES REPRODUCTION ANIMAL see also ANIMAL SYMMETRY [E H A]

Bibliography H K Harring Synopsis of the Rotatoria, *U S Natl Museum Bull* 81 1913 C T Hudson and P H Gosse *The Rotifera or Wheel animalcules Both British and Foreign* 2 vols 1886–1889 F J Myers Rotifer fauna of Wisconsin V *Trans Wisconsin Acad Sci* 25 353–411 1930

Rous sarcoma

A virus which produces a disease in chicks pheasants and ducklings

The Rous sarcoma virus which was discovered in 1911 has been used as the standard example of a

lated into newborn chicks or into the embryo, it causes a fatal hemorrhagic disease It is probably part of the leukosis group of fowl diseases which may be caused by a single virus See ANIMAL VIRUS, AVIAN LEUKOSIS, TUMOR VIRUSES [A E M]

Rubber

Until recently the word rubber connoted natural or tree" rubber which is a hydrocarbon polymer of isoprene units The more recently developed synthetic rubbers possess different chemical structures but resemble tree rubber in many physical properties The most recent development is a synthetic natural rubber that is a duplicate of the natural rubber molecule

The industrial importance of synthetic rubbers can hardly be overemphasized Some types have invaded the fields of tires hose and belts for which

ble

Structural prerequisites of both natural and synthetic rubbers are long threadlike molecules Their characteristic property of reversible extensibility results from the randomly coiled arrangement of the long polymer chains Upon extension the chains are distorted but like a spring they revert to the kinked arrangement upon removal of the stress

Smoked sheet and pale crepe represent the forms in which the major portion of natural rubber is commercially available The smoked sheet is obtained by acid coagulation of the tree latex sheeting the coagulum and then drying and smoking the

or after coagulation with sodium bisulfite and then washing drying and sheeting Synthetic rubbers are usually available as baled sheets made from the coagulum of a polymerization process

Latex a colloidal suspension of polymers in an aqueous medium is another form in which both natural and synthetic rubbers are commercially available The various types of latex include the product of the rubber tree the suspensions resulting from the emulsion polymerization of various monomers and the dispersion of the bulk dry polymer in an aqueous medium Although foamed products represent the largest field of application for natural rubber latex synthetic latexes are being used at a rapidly growing rate for this purpose Latex is also used to impregnate textiles cords and paper to coat fabrics and to make adhesives, water based paints and articles molded by casting methods

PROCESSING

In the crude state natural and synthetic rubbers possess certain physical properties which must be modified to obtain useful end products The natural forms are weak and adhesive They lose their elas

cretory system consists primarily of paired nephridial tubes which are provided with flame bulbs These open posteriorly either into a bladder, or directly into the cloaca The female reproductive organs in most rotifers consist of a single ovary and a vitellarium which open into the cloaca however, exceptions to this occur in the Seisonacea and Bdelloidea The foot or tail contains cement glands whose secretion enables many species to attach themselves firmly to desired substrata Sensory organs consist of one or more eyespots which may be lacking in some species as well as sensory bristles and papillae

Rotifers are covered by a nonchitinous cuticle which may be divided into annuli especially in Bdelloidea and Notommatidae and which in some rotifers is thickened into a lorica Loricate rotifers often have characteristic patterns and ornamentations which are useful in their systematics Sessile rotifers are generally provided with a tube often gelatinous but in some rotifers especially *Floscularia*, this is quite intricate in construction

There is a fairly complex system of musculature It includes bands of circular muscles in the body wall which allow the rotifer to elongate longitudinal muscles in the body wall which control the expansion and retraction of the head and foot muscles associated with the viscera and others

Rotifers lack definite cells in most parts of their bodies hence a syncytial condition prevails Nevertheless there is a constancy in the number of nuclei present in different organs The common rotifer *Epiphanes* (= *Hydatina*) *senta* widely used as an experimental animal is reported to have approximately 960 cells (or nuclei) in its body

The pharynx is modified into a structure peculiar to rotifers namely a muscular mastax which contains a masticatory apparatus the trophus The terms mastax and trophus are used as synonyms by many workers Typical trophi contain 7 parts 3 central pieces comprising the incus and 4 lateral pieces forming the paired mallei (no relation to bones of the middle ear of mammals) The 3 parts of the incus include a posterior supporting structure the fulcrum and 2 rami Each malleus consists of 2 parts a head or uncus and a handlelike piece or manubrium Some trophi contain additional accessory parts The 8 primary types of trophi found in rotifers are as follows fulcrate (Seisonacea) (Fig 1) ramate (Bdelloidea) uncinate (Collothecacea) malleate (brachionids) virgate (notommatids) incudate (asplanchnids) forcipate (dicranophorids) and cardate (*Lindia*) There are various intermediate types such as malleoramate (Flosculariacea) and virgo-forcipate (*Synchaeta*)

In rotifers that feed exclusively on microplankton and particulate matter the rami are developed for grinding (ramate and malleoramate) (Fig 2) Predatory rotifers have protuberant trophi used for grasping and are of the incudate and forcipate types Less specialized trophi such as malleate virgate cardate and uncinate (Fig 3) occur in rotifers that are incidentally predatory but that feed primarily on algae and particulate matter

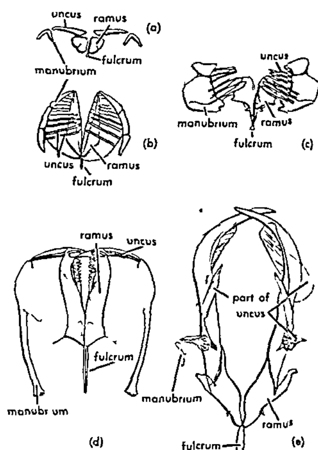


Fig 2 (a) Ramate trophus of a bdelloid rotifer (schematic) ventral view (b) Upper view (After Bartos) (c) Malleoramate trophus of *Hexarthra* upper view (After Hauer) (d) Forcipate trophus of *Dicranophorus* ventral view (After Hauer) (e) Incudate trophus of *Asplanchna* ventral view (After de Beauchamp)

Reproduction The most complex mode of reproduction in rotifers is found in the order Monogononta The sexes are markedly dimorphic and males produced only sporadically The usual mode of reproduction is by parthenogenesis. Females produce diploid amictic eggs which develop without fertilization into females At intervals mictic eggs are produced which have undergone maturation and are haploid If unfertilized a mictic egg develops into a male rotifer hence males are haploid If fertilized the mictic egg becomes a thick shelled resting egg which can remain dormant for weeks or months but which eventually develops into a female Asexual (parthenogenetic) reproduction does not occur among the Seisonacea in which both males and females are diploid In the Bdelloidea on the other hand only parthenogenetic reproduction is known

Distribution. The rotifers are among the most widely distributed of aquatic animals They occur in all types of fresh water habitats from large lakes to temporary ponds and mud holes The genus *Keratella* is the most common metazoan animal in the plankton of lakes and ponds other rotifers occur abundantly in the littoral zone and on aquatic vegetation Frank J Myers, who was the outstand

nols The choice of the antioxidant is governed by t · · · · · · · ·purpose for which it is intended
·
ends of the polymer chain and preven · chain degradation of the rubber

Although most antioxidants have staining characteristics, this property does not interfere with their use in rubber compounds for tires hose, and belts However light colored end products or rubber compounds which will come in contact with light-colored surfaces necessitate the use of non staining antioxidants These compounds are usually highly 'hindered' phenols obtained by alkylation of phenols or cresols or are derivatives of aromatic phosphite esters See ANTIOXIDANT

The cracking of rubber compounds under stress when exposed to very low ozone concentrations such as exist in the atmosphere also originates from the unsaturated chemical structure of the rubber Although the deteriorating effects of ozone can be · · ·

tain accelerators have antiozonant properties but the majority of the commercial antiozonants are substituted p phenylenediamines These can be subdivided into the diaryl aryl alkyl and diaryl p phenylenediamines Another type is formed by the condensation of amines and ketones

anism of antioxidants The protective effect is believed to be the result of a preferential combination of the ozone with the antiozonant instead of · · ·

appear · ·
ultraviolet and the blue end of the spectrum is not reduced to any appreciable extent by antioxidants Although complete protection against photooxidation · · ·
suff · · ·
vide a certain degree of protection for light colored products

At the present time the effect of γ-radiation upon the various rubbers is very complicated and unpredictable with respect to its details The changes in the physical properties of the rubbers are governed by whether the primary results of the exposure are cross linking chain scission of the polymer molecules · · · In turn these molecular changes are dependent upon the molecular structure of the rubber radiation dosage and exposure conditions which determine whether such physical properties as tensile strength or elongation decrease increase or remain the same The changes induced by radiation are the result of a free radical mechanism, and the type of compounding can affect the results.

Two methods are available to combat the degradative effects of radiation exposure of rubbers One of these is the development of new synthetic rubbers with inherent radiation resistant characteristics. For example, the recently introduced adduct rubbers exhibit exceptional stability against the combined effects of heat and radiation Another · · · the incorporation of protective agents

tiveness as radiat · · ·
have shown that the following compounds impart radiation stability to various rubbers N,N' cyclohexylphenyl p phenylenediamine a mixture of diphenyl p phenylenediamine and phenyl α naphthyl amine, quinhydrone, diphenylamine and p-meth · · · of the

PHYSICAL TESTING

The physical testing of rubber compounds plays an essential role in rubber technology Because the rubbers available today exhibit a wide range of properties which can be further varied by compounding techniques, standard methods of evaluation based upon physical properties have been established

Physical tests screen measure, and evaluate the desired physical properties of a product according to the information requested In the rubber industry tests are conducted on natural and synthetic rubbers rubberlike materials, plastics, fabrics and cords Current advancements into numerous fields have expanded testing procedures to include almost any physical property of almost any material that is capable of being tested

Standards of the American Society for Testing Materials (ASTM) are used to test materials whenever possible This standardization of testing procedures simplifies the demands of the manufacturer and the compounder, and guarantees universal comparable data

In screening for physical properties, the first step in evaluation of unvulcanized compounds is to measure their relative plasticity or viscosity This is done by means of the Mooney Plastometer or more recently by the Flow Tester

Stress-strain properties are of importance when the final product is to be used under dynamic conditions These properties include stress, which is the intensity of internal forces which act on a given plane through a point, strain which is a measure of change in size or shape due to force, modulus, the ratio of stress to corresponding strain below the proportional limit, elongation the increase in the original length of that portion of the specimen over which strain or change of length is desired This is usually measured after fracture in tension of the specimen

Stress-strain properties are evaluated by the use of a constant power driven machine such as the Scott Tensile Tester With the advancement in

ticity by *constant* use and are susceptible to temperature variations. Consequently, it is necessary to process the crude rubbers by compounding and vulcanization procedures which *transform them* into products which can fulfill a specific function.

Curatives and vulcanization. After the addition of curing or vulcanizing agents to rubber, the application of heat causes a chemical reaction to take place to give a durable product. Sulfur, which was the first successful curing ingredient, is still the basis of the vulcanization system used in industry today. However, various chemicals or combinations of chemicals are also capable of vulcanizing rubber. These include oxidizing agents such as selenium, tellurium, organic peroxides, and nitro compounds, and also generators of free radicals such as organic peroxides, azo compounds, and certain organic sulfur compounds such as the alkyl thiuram disulfides.

Research has resulted in the introduction of several classes of compounds which accelerate the rate of vulcanization. These accelerators possess the additional desirable feature of reducing the sulfur requirements and in general enhance the physical properties of the vulcanized rubber.

Inorganic chemicals, such as the basic carbonates and oxides of lead supplemented by magnesia or lime, were the first accelerators to be used in the rubber industry. Since the introduction of the first organic accelerator in 1906, several thousand rubber accelerators have been patented, but only a small number of these are in general use. Most of the accelerators in current use can be grouped into the following general classifications: (1) aldehyde amines, (2) guanidines, (3) thiuram sulfides, (4) thiazoles, (5) thiazoline, (6) dithiocarbamates, and (7) mercaptoimidazolines.

Rubber can be vulcanized by gamma radiation, but it is not economically feasible at the present time. Items as large as a tire have been vulcanized by this means and have shown wearing and aging properties comparable to tires vulcanized by the ordinary procedures.

Not all rubbers can be vulcanized by irradiation. Butyl rubber is an example of a type which undergoes degradation rather than cross-linkage. The structure of the chain molecule is the determining factor. Cross-linking predominates if the polymer has the structure $(-CH_2CH_2-)_n$ or $(-CH_2CHR-)_n$. If the structure is $(-CH_2CR_1R_2-)_n$, degradation is the rule. There is great variation in the effects of γ radiation vulcanization, depending upon the type of rubber being treated, the nature of the compounding ingredients, and the conditions. Although the vulcanization process was discovered by Charles Goodyear in 1839, there is still no completely satisfactory explanation of the mechanism of the vulcanization procedure or the accelerator action. However, it is agreed that any vulcanization procedure involves a cross-linkage of the long chain polymer molecules by means of sulfur bridges or ionic linkages to form a network structure. This type of structure reduces the essentially thermoplastic properties of the crude rubber and confers predominantly elastic properties to the resulting vulcanizate by preventing slippage of the long chain molecules past each other. A possible explanation of the mechanism of accelerator action is the generation of free radicals by the specific accelerators under vulcanization conditions.

Pigments. Vulcanization improves the elasticity and aging properties of crude rubber, but in most cases it is necessary to further enhance such properties as tensile strength, abrasion resistance, and tear resistance by the incorporation of fillers. The fillers which improve specific physical properties are known as reinforcing fillers; those which serve primarily as a diluent are classed as inert fillers. The physical properties of the resulting vulcanizate are affected by both the type and the amount of filler.

Carbon black is the most universally used filler in the rubber industry. The three types of carbon blacks used commercially in the greatest bulk for this purpose include the channel, furnace, and thermal blacks. Each of these types may be further classified according to particle size and surface structure and then selected according to the specific properties which are required in the end product. See CARBON BLACK.

In addition to the carbon blacks, inorganic reinforcing agents, such as zinc oxide and the silicas, are used for the reinforcement of light colored end products. Although zinc oxide is not used extensively in the rubber industry today, its incorporation does enable the resulting product to withstand extended exposure at high temperatures, and it also functions as an activator during the vulcanization process. The silicas are used in those products in which high abrasion resistance is an essential requirement.

The inert fillers such as whiting and various types of clays serve merely as extenders and to facilitate the processing of the compound. In general they do not improve the tensile strength of the resulting vulcanizate.

Protective agents. The aging stability of rubber compounds is influenced by such factors as heat, light, and atmospheric conditions. The chemical unsaturation of natural and synthetic rubber compounds provides a focal point for oxidation. The initial oxidative degradation involves breaking of the polymer chains and cross-linking, which results in loss of tensile strength and stiffening of the vulcanizate. Depending upon the conditions, the free radicals produced during chain scission are subject to further oxidation with the formation of lower molecular species. The extent and nature of degradation are dependent upon the specific rubber and the conditions of the exposure.

Improvements in the resistance to attack by oxygen have been accomplished by the incorporation of chemical compounds, known as antioxidants, into the rubber compound.

Broadly speaking, antioxidants may be classified into two general groups: aromatic amines and phe-

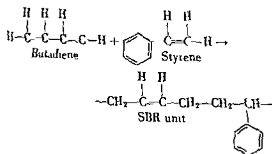
ination costs and 60% in general costs In addition to this reduced cost of production the higher yield ing clones would result in greater potential pro duction of rubber

The application of chemicals such as copper sul fate 2 4-D and 2,4,5 T to the bark of the tree has greatly increased rubber yield The use of these chemicals has to be carefully controlled because when used in excess they may either inhibit bark renewal or make the tree more susceptible to dis ease For example a small amount of cobalt has a beneficial effect on the growth of the rubber plant but larger concentrations may depress growth Re cent data indicate that when the proper soil chem istry is maintained the tree will develop maximum rubber yield with a minimum of stimulation

Butadiene styrene rubbers (GR S SBR) The extensive development of the synthetic rubber in dustry originated with the World War II emer gency but the continued expansion has been the result of the superiority of the various synthetic rubbers in certain properties and applications Sta tistics show that in 1957 synthetic rubbers repre sented 40% of the total world consumption and 63% of the total rubber consumption in the United States

The butadiene styrene rubber formerly desig nated as GR S types but now called SBR consti t tes the b lk of the synthetic rubber production

type the ratio of butadiene to styrene is approxi mately 78 22



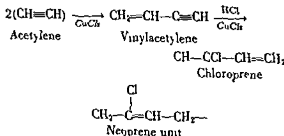
The original or hot SBR is polymerized at a tem perature of 122°F or higher and is used primarily for mechanical goods because of its better color re tention The cold type polymerized at 41°F ex hibits an improvement over the hot type in most physical properties When used in light duty tires such as passenger car tires the cold rubber com pounds have proved equal or superior to natural rubber treads However they are inferior to natural rubber for truck tires because of the greater heat buildup during flexing

The cold oil extended type prepared by the re placement of a portion of the polymer by a heavy fraction oil accounts for approximately 25% of the cold rubber production Its advantage is primarily economic Master batches containing both oil and carbon black have the advantage of process sim plification.

In general the compounding and processing meth ods used for all the SBR types are similar to those of natural rubber Although natural rubber is superior with respect to lower heat buildup, re silience and hot tear strength the SBR types are more resistant to abrasion and weathering How ever carbon
must be ad
develop the

In addition to tires other end use items include belting hose molded goods shoe soles flooring, and insulation

Neoprene Neoprene one of the first synthetic rubbers used commercially in the rubber industry is a polymer of chloroprene 2 chlorobutadiene 1 3 In the manufacturing process acetylene the basic raw material is dimerized to vinylacetylene and then hydrochlorinated to the chloroprene monomer



An emulsion system is used for the polymeriza tion and the resulting polymer is isolated by freeze coagulation Copolymer types of neoprene in which the chloroprene monomer predominates are also available Other commercial types of neo prene are manufactured which differ according to compounding variations dependent upon end usage

Sulfur is used to vulcanize some types of neo prene but most of the neoprenes are vulcanized by the addition of basic oxides such as magnesium oxide and zinc oxide Other compounding and proc essing techniques follow similar procedures and use the same equipment as for natural rubber One of the outstanding characteristics of neoprene is the good tensile property without the addition of carbon black filler However carbon black and other fillers can be used when reinforcement is re quired for specific end use applications requiring increased tear and abrasion resistance

Because the neoprenes have outstanding resist ance to ozone weathering a variety of chemicals oil and flame they have been used for applica tions where natural rubber is inadequate A possi ble limitation of more extensive applications of the neoprenes is their relatively high freezing points which reduce the serviceable temperature range Another disadvantage is the cost factor which limits the applications to items which can justify premium prices

Typical uses of the various neoprenes include oil resistant applications hose low voltage insula tion adhesives molded goods and tank linings

electronics, the Instron Tester, and similar apparatus employing load cells, has come into practical use for obtaining more detailed information regarding such physical testing

All physical testing is conducted at a standard temperature of 77°F and 50% relative humidity, except where higher or lower temperature testing is requested

Accelerated aging and conditioning of materials prior to testing is conducted in specially designed and automatically controlled ovens and refrigerators Physical tests may be requested for properties such as hardness or softness tear resistance abrasion, flex, resistance to ozone, ultraviolet, salt spray and water, compression and recovery permeability to gases, volatility, rebound, elasticity dynamic modulus and resilience, and low temperature properties Each specific test employs a specially designed apparatus for measuring the physical property or properties desired

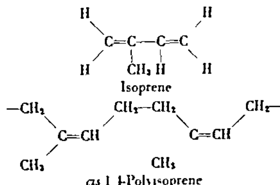
An example of a modern piece of test equipment is the Gehman Torsion Apparatus It has automatic controlling devices whereby measurements can be made at any desired temperature from ambient to −320°F by use of liquid gases such as nitrogen

Any change in formulation, compounding or processing of a material such as rubber is quite discernible in the physical testing data, thereby proving the importance of such a procedure either in research development or in production Sometimes a slight deviation or the addition of a foreign element may cause an inferior product Frequent physical testing helps to ensure a standard product

SPECIFIC TYPES OF RUBBER

There is only one rubber obtained from natural sources in commercial quantities There are several synthetically produced compounds that are classed as rubbers *See* POLYMER PROPERTIES

Natural rubber. Although natural rubber may be obtained from hundreds of different plant species the most important source is the rubber tree *Hevea brasiliensis* Natural rubber is *cis* 1 4-polyisoprene containing approximately 5000 isoprene units in the polymer chain



The world production of natural rubber has increased from 45 000 long tons in 1900 to 1 902 500 long tons in 1957, a 42 fold increase Added to the 1957 figure is 1 261 916 long tons of synthetic rubbers making a grand total of all rubber production in 1957 of 3,164 416 long tons a 70-fold increase since 1900

During the past several years the emphasis in research and development of natural rubber at the various rubber institutes and at the larger and more progressive plantations has shifted from chemistry to botany The major current chemical investigation is directly allied with botanical pursuits, and has the ultimate aim of increasing the rubber yield of the trees This change in emphasis has been the result of the tremendous economic challenge of synthetic rubber for the world markets Because natural rubber producers had experienced the loss of the bulk of the market in the United States because of lower priced synthetic, they were faced with the realization that as soon as the European synthetic rubber plants were in production, natural rubber consumption would be considerably reduced Consequently yield increase and price reduction were necessary

The cross pollination of high rubber yielding clones of *H brasiliensis* has resulted in a significant increase in rubber production The Rubber Research Institute of Malaya reported that two of their "Series 600" clones yielded in excess of 920 lb/(acre)(year) during the first year (1956) of commercial tapping and 6–7 years after budding This is more than double the yield of average plantation stock Several experimental clones have produced in excess of 2000 lb/(acre)(year)

It has been estimated by the Government of the Federation of Malaya that the replacement of the old trees with modern high yield trees would result in a decrease of 35% in tapping costs 65% in cul

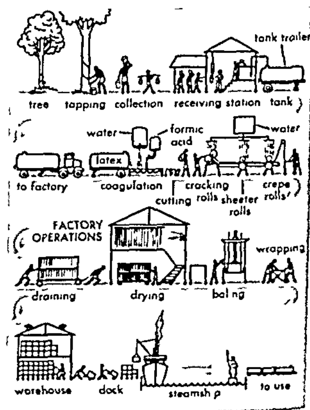


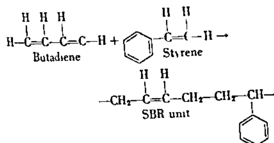
Fig 1 Steps in natural rubber production

tn
to t s
ing clones would result in greater pote
duction of rubber

The application of chemicals such as copper sulfate 2,4-D and 2,4,5-T to the bark of the tree has greatly increased rubber yield. The use of these chemicals has to be carefully controlled because, when used in excess, they may either inhibit bark renewal or make the tree more susceptible to disease. For example, a small amount of cobalt has a beneficial effect on the growth of the rubber plant ... may depress growth. Re-

Butadiene styrene rubbers (GR S, SBR) ... extensive development of the synthetic rubber industry originated with the World War II emergency but the continued expansion has been the result of the superiority of the various synthetic rubbers in certain properties and applications. Statistics show that in 1957 synthetic rubbers represented 40% of the total world consumption and 63% of the total rubber consumption in the United States.

The butadiene-styrene rubber, formerly designated as GR S types but now called SBR, constitutes the bulk of the synthetic rubber production. All the SBR type rubbers are obtained by the emulsion polymerization of butadiene and styrene in varying ratios. However, in the most commonly used type the ratio of butadiene to styrene is approximately 78:22.



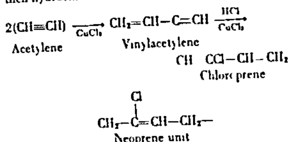
The original or hot SBR is polymerized at a temperature of 122°F or higher and is used primarily for mechanical goods because of its better color retention. The cold type, polymerized at 41°F, exhibits an improvement over the hot type in most physical properties. When used in light duty tires such as passenger car tires, the cold rubber compounds have proved equal or superior to natural rubber treads. However, they are inferior to natural rubber for truck tires because of the greater heat buildup during flexing.

The cold oil extended type, prepared by the replacement of a portion of the polymer by a heavy fraction oil, accounts for approximately 25% of the cold rubber production. Its advantage is primarily economic. Master batches containing both oil and carbon black have the advantage of process simplification.

In general the compounding and processing methods used for all the SBR types are similar to those of natural rubber. Although natural rubber is superior with respect to lower heat buildup, resilience, and hot tear strength, the SBR types are more resistant to abrasion and weathering. However, carbon black or some other reinforcing fillers must be added to the butadiene-styrene rubber to develop the best physical qualities.

In addition to tires, other end use items include belting, hose, molded goods, shoe soles, flooring and insulation.

Neoprene. Neoprene, one of the first synthetic rubbers used commercially in the rubber industry, is a polymer of chloroprene, 2-chlorobutadiene-1,3. In the manufacturing process, acetylene, the basic raw material, is dimerized to vinylacetylene and then hydrochlorinated to the chloroprene monomer.



An emulsion system is used for the polymerization, and the resulting polymer is isolated by freeze coagulation. Copolymer types of neoprene, in which the chloroprene monomer predominates, are also available. Other commercial types of neoprene are manufactured which differ according to compounding variations dependent upon end usage.

Sulfur is used to vulcanize some types of neoprene, but most of the neoprenes are vulcanized by the addition of basic oxides such as magnesium oxide and zinc oxide. Other compounding and processing techniques follow similar procedures and use the same equipment as for natural rubber. One of the outstanding characteristics of neoprene is the good tensile property without the addition of carbon black filler. However, carbon black and other fillers can be used when reinforcement is required for specific end use applications requiring increased tear and abrasion resistance.

Because the neoprenes have outstanding resistance to ozone, weathering, a variety of chemicals, oil, and flame, they have been used for applications where natural rubber is inadequate. A possible limitation of more extensive applications of the neoprenes is their relatively high freezing points, which reduce the serviceable temperature range. Another disadvantage is the cost factor, which limits the applications to items which can justify premium prices.

Typical uses of the various neoprenes include oil resistant applications, hose, low voltage insulation, adhesives, molded goods, and tank linings.

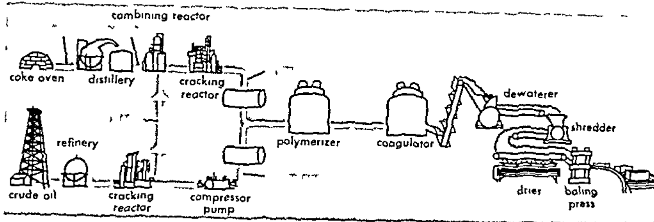
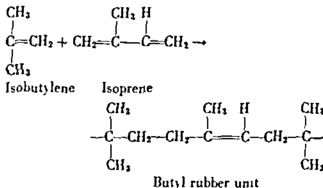


Fig 2 Steps in production of one type of synthetic rubber (butadiene styrene rubber) (*Firestone Tire and Rubber Co*)

Butyl rubber. Isobutylene and isoprene or butadiene obtained from cracked refinery gases are the primary raw materials required for the manufacture of Butyl rubbers The small ratio of diolefin varying from 0 5 to 4 5%, provides the unsaturation necessary for vulcanization

A solution polymerization system is used the reaction being conducted at approximately −150°F in the presence of a dilute solution of aluminum chloride $AlCl_3$ in methyl chloride as a catalyst



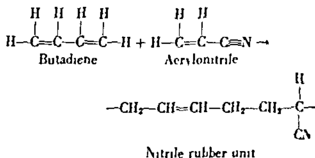
Butyl rubbers are processed according to the conventional methods used for natural rubber and SBR An antioxidant sulfur and an ultra accelerator are compounded into the stock However because of the low degree of unsaturation in the Butyl rubber molecule it is more difficult to complete the cure, and consequently the vulcanization requires higher temperatures and longer times

Neither neoprene nor Butyl rubber requires carbon black to increase their tensile strengths, but the reinforcement of Butyl rubber by carbon black or other fillers does improve the modulus and increases the resistance to tear and abrasion

The excellent resistance of Butyl rubbers to oxygen, ozone, and weathering can be attributed to the small amount of unsaturation present in the polymer molecule In addition, these rubbers exhibit good electrical properties and high impermeability to gases

Tires made of Butyl rubber are now in the development stage Specialty applications include coating compositions, tank linings, adhesives, and bags for the bag molding of plastics However, the bulk of the Butyl rubber production has been used for tire inner tubes because of its outstanding air retention characteristics Another large outlet for Butyl rubber stocks is wire and cable insulation

Nitrile rubber. Much of the basic pioneering research on emulsion polymerization systems was with the nitrile-type rubbers These rubbers which originated as the German buna N types in 1930 are copolymers of acrylonitrile and a diene usually butadiene Hydrogen cyanide and ethylene oxide are the raw materials currently used for the production of acrylonitrile



The ratios of the monomers can vary from 80 20 to 55/45 butadiene/acrylonitrile The oil resistant properties of the rubber increase with increasing nitrile content but there is a corresponding decrease of the low temperature flexibility of the rubber

Both sulfur and nonsulfur vulcanizing agents may be used to cure these rubbers Carbon black or other reinforcing agents are necessary to obtain the optimum properties

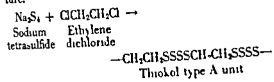
If proper processing methods are followed the nitrile rubbers can be blended with natural rubber polysulfide rubbers and various resins to provide specific characteristics in the resulting blend, such as increased tensile better solvent resistance, and improved weathering resistance

Nitrile rubbers are also called buna N Chemigum N Hycar OR Perbunan and Butaprene N

Excellent oil grease and solvent resistance are the outstanding properties of the nitrile rubbers Consequently, the commercial usage of these rub-

bers is largely for items in which these properties are essential. Another major usage is the utilization of the latex form for adhesives and for the finishing of leather, impregnation of paper, and the manufacture of nonwoven fabrics.

Thiokol polysulfide rubber. Polysulfide rubbers obtained by the reaction of an organic dihalide and an inorganic polysulfide, are condensation polymers of varying viscosities. The original Thiokol polysulfide rubber was the reaction product of ethylene dichloride and sodium tetrasulfide. The resulting polymer is believed to be linear in structure.



$Na_2S_4 + ClCH_2CH_2Cl \rightarrow$ —$CH_2CH_2SSSSCH_2CH_2SSSS$—

Sodium tetrasulfide, Ethylene dichloride, Thiokol type A unit

By conducting the reaction in an aqueous me- l

free from sodium chloride and then dried.

The use of different dihalides and different sul . . . l properties . . .

number of sulfur atoms in the polymer chain . . addition to ethylene dichloride the most generally used dihalides are methylene chloride, propylene dichloride, glycerol dichlorohydrin, dichloroethyl ether, dichloroethyl formal, and triglycol dichloride.

Sulfur does not function as a vulcanizing agent for these rubbers. However the necessary crosslinks between the polymer chains can be attained . . .

. . . porate reinforcing pigments such as carbon black to improve tensile strength, hardness, and abrasion resistance. Other fillers such as zinc oxide, zinc sulfide, titanium dioxide, and lithopone are used for nonblack stocks.

The recently introduced polysulfide liquid polymers obtained by controlled reduction of high molecular weight polymers are finding applications as sealants, flexible molds and patterns, epoxy resin modifiers, and leather impregnants.

The outstanding properties of these rubbers are their high solvent and oil resistance, their excellent resistance to aging, ozone, and sunlight, and their very low permeability to gases and vapors. These characteristics make them useful for such applications as gas hose for cars and service stations, printers' blankets, cable coverings, and coating for balloon fabrics. More extensive application is limited because the tensile strength and abrasion resistance are lower than many of the other synthetic rubbers. The disagreeable odor of some of the types is another restriction for certain applications.

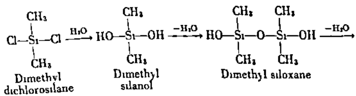
Silicone rubbers. Silicone rubber, like the Thiokols, is a linear condensation polymer based on a . . . aration di . . .

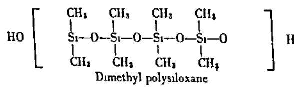
siloxane, and this, upon further condensation yields dimethyl polysiloxane, the standard silicone rubber as shown below.

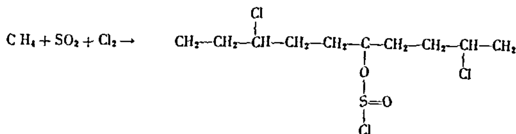
Various types of silicone rubbers are produced by substi . . . polymer . . . groups . . . evidenced by improvements in specific properties. For example, the presence of phenyl groups in the polymer chain gives further improvement in low temperature properties.

Because sulfur is not effective for the vulcanization of most silicone rubbers, a strong oxidizing agent such as a benzoyl peroxide is used; the crosslinking produced is random. However, when unsaturated groups such as vinyl or allyl are present these products are vulcanized with sulfur, the vinyl groups serving as a control of the degree of cross linking.

Although the standard silicone rubbers are not reinforced by carbon black, the physical properties can be improved by the incorporation of various inorganic fillers such as titania, zinc oxide, iron oxide, and silica, which act both as reinforcing and modifying agents. The physical, chemical, and electrical properties can be altered by varying the type and amount of these fillers. Carbon black can be used as a filler with vinyl containing polymers.







Hypalon unit

In general the silicone rubbers have relatively poor physical properties and are difficult to process However they are the most stable of rubbers and are capable of remaining flexible over a temperature range of $-130°$ *to* $+600°F$ *They are unaffected* by ozone are resistant to hot oils and have excellent electrical properties. *Their most extensive* applications are for wire and cable insulation tubing packings and gaskets In the form of dispersions *and pastes they* are used *for* dip coating spraying brushing and spreading *See* SILICON

Hypalon Hypalon a *chlorosulfonated* polyethylene is prepared by treating polyethylene with a *mixture of chlorine* and sulfur dioxide whereby a few scattered chlorine sulfonyl chloride groups are introduced into the polyethylene chain By this treatment polyethylene is converted into a rubber *like material* in which the undesirable degree of crystallinity is destroyed but the desirable proper*ties of polyethylene* still maintained

The ratio of the chlorine and sulfur atoms is care*fully controlled* so that there is approximately 29% chlorine and about 1 25% sulfur The outstanding chemical stability of Hypalon results from the complete absence of unsaturation in the polymer chain

Vulcanization is accomplished by means of metallic oxides such as litharge magnesia or red *lead* in the presence of an accelerator The sulfonyl chloride groups provide the sites of reactivity at which the bases react with the chlorine of the sulfonyl chloride group to give cross links The incorporation of an organic acid such as hydrogenated wood rosin or stearic acid is also necessary for optimum cure Carbon black or other fillers are not needed to obtain optimum strength properties and antioxidants are required only for maximum resistance to heat In addition to its use as a stock in itself Hypalon can be blended with other types of rubbers to provide a wide range of properties

Extreme resistance to ozone is the most important property of Hypalon Its chemical resistance to strong chemicals such as nitric acid sulfuric acid chromic acid hydrogen peroxide and strong bleaching agents is superior to any of the commonly used rubbers These vulcanizates also have good heat resistance mechanical properties and unlimited colorability

Typical applications include covers for conveyor belts steam hose tubes O rings and gaskets in ozone generators miscellaneous molded goods and coated fabrics for outdoor service

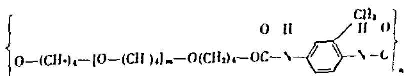
Polyurethane rubbers Polyurethane rubber is an all inclusive name given to elastic polymers con*taining the* urethane *linkage* In general they are prepared by extending a polyester a polyalkylene ether glycol or a polyhydrocarbon diol with diisocyanates the extension of bis chloroformates of such polymeric diols with diamines is also a practical method of rubber synthesis In either case, the chosen diol is usually in the 1000–5000 molecular weight range.

The most common polyesters are adipates of C_4 C_3 or $C_?$ glycols whereas polybutylene ether glycol from the cationic polymerization of tetrahydrofuran dominates the polyether field Little has been reported on the hydrocarbon diols with exception of polychloroprene diols

A basic repeating structure for a typical urethane rubber can be graphically depicted as shown below

The cross linking reactions for converting polyurethane rubbers to useful products depend upon the particular type of processing steps desired The so called millable or processible rubbers can first be fully compounded and then subsequently milled extruded calendered and molded much like the diene rubbers The actual vulcanization reactions are commonly conducted at temperatures below 300°F Cross linking was first achieved by diisocyanate reactions with active hydrogens in the polyurethane rubber chain Current practice is to prepare the polyurethane chain with reactive centers such as olefin bonds so that the more conventional vulcanization by sulfur is possible

The casting polyurethane rubbers are processed differently from the millable rubber and are distinctively characterized by a relatively short pot life after being fully compounded In a typical case a polyester is pretreated with a molecular excess of di or polyisocyanate The resulting prepolymer in a molten condition is mixed with a reticulating agent such as an aromatic diamine and quickly poured into a mold Chain extension and cross-linking occur concurrently in casting systems

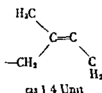


Exten ive physical data have been reported only on polyesterurethanes and polyetherurethanes. These rubbers excel in strength resilience, and resistances to al rasion ultraviolet light oxygen ozone solvents, and flex cracking They are satisfactory but not superior at extreme temperatures

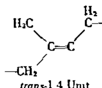
Particularly interesting characteristics are their use as a replacement for metals in many moving part applications and as foams.

cis 1,4 Polyisoprene. In the last few years synthetic *cis* 1 4-polyisoprene has been made from isoprene by the use of two different classes of catalysts. The first class includes lithium metal and the lithium alkyls The second class uses a mixture of an aluminum alkyl and titanium tetrachloride, the system first used for the low pressure polymerization of ethylene by Karl Ziegler Both catalyst systems are carried out in hydrocarbon solution and require highly purified monomer and solvent Traces of air moisture and most polar compounds adversely affect reaction rates polymer properties, and structure

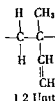
The *cis* 1 4 polymer structure obtained with these catalysts is also characteristic of natural *Hevea* rubber The presence of a high cis content appears necessary for the desirable physical properties obtained with *Hevea* in contrast with the inferior properties of emulsion type polyisoprene which contains mixed *cis*, *trans* 1 2 and 3 4-isoprene units in the polymer chain



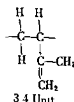
cis 1 4 Unit



trans-1 4 Unit



1 2 Unit



3 4 Unit

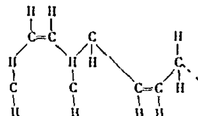
When the polymerization is complete the catalyst is destroyed for example by treating with alcohol and the solvent removed The solid polymer thus obtained may then be handled and compounded as any other dry rubber is

In spite of the general similarity of the polymers obtained with the two catalyst systems there are

of crystallization The lithium system gives polymers of extremely high molecular weight, even higher than natural rubber, whereas the Al Ti rubbers are generally lower in molecular weight than *Hevea*

Both types appear to be substantially equivalent to *Hevea* for tire applications The lithium rubbers give somewhat better resilience properties than the Al Ti polymers but these in turn generally have better hot tensile strengths Further developments of the next few years will undoubtedly close any performance gaps that remain between the synthetic and natural products

cis 1,4 Polybutadiene In 1956 and 1957 two new catalyst systems were disclosed for preparing high (> 90%) *cis* 1 4-polybutadiene



One system employs a modification of the conventional Ziegler catalyst in which titanium tetra iodide is treated with trialkyl aluminum to form the active catalyst The Ziegler catalyst gives a rubber containing a lower proportion of *cis* 1 4-polybutadienes The other system employs as catalyst a modified Natta stereospecific catalyst complex obtained from the reaction of butyl lithium with titanium trichloride Both systems employ solution polymerization with hydrocarbon diluent Other catalyst systems disclosed for polybutadiene em

indicate lower heat generation, higher resilience as well as improved low temperature properties when compared with butadiene styrene emulsion prepared polymers The low temperature behavior of the polymer is an improvement over natural rubber Tensile strength and elongation are not equivalent to natural rubber The processing characteristics of this polymer are inferior to those of natural rubber

The use of 1 1 blends of 95% [illegible]

[illegible] ce in experimental retread tires A similar blend employed in truck tires has given indications of better abrasion resistance and measurably lower heat build than natural rubber Improved processing of the polybutadiene polymer is obtained by employing lower molecular weight polymer

Adduct rubbers The adduct rubbers represent one of the most recent additions to the family of

$$C_2H_4 + SO_2 + Cl_2 \rightarrow CH_2{-}CH_2{-}CH(Cl){-}CH_2{-}CH_2{-}C(O{-}S(=O){-}Cl){-}CH_2{-}CH_2{-}CH(Cl){-}CH_2$$

Hypalon unit

In general, the silicone rubbers have relatively poor physical properties and are difficult to process However they are the most stable of rubbers, and are capable of remaining flexible over a temperature range of −130° to +600°F They are unaffected by ozone, are resistant to hot oils, and have excellent electrical properties Their most extensive applications are for wire and cable insulation tubing, packings, and gaskets In the form of dispersions and pastes, they are used for dip coating, spraying, brushing, and spreading *See* SILICON

Hypalon. Hypalon, a chlorosulfonated polyethylene, is prepared by treating polyethylene with a mixture of chlorine and sulfur dioxide, whereby a few scattered chlorine sulfonyl chloride groups are introduced into the polyethylene chain By this treatment polyethylene is converted into a rubber like material in which the undesirable degree of crystallinity is destroyed but the desirable properties of polyethylene still maintained

The ratio of the chlorine and sulfur atoms is carefully controlled so that there is approximately 29% chlorine and about 1 25% sulfur The outstanding chemical stability of Hypalon results from the complete absence of unsaturation in the polymer chain

Vulcanization is accomplished by means of metallic oxides such as litharge magnesia or red lead in the presence of an accelerator The sulfonyl chloride groups provide the sites of reactivity at which the bases react with the chlorine of the sulfonyl chloride group to give cross links The incorporation of an organic acid such as hydrogenated wood rosin or stearic acid is also necessary for optimum cure Carbon black or other fillers are not needed to obtain optimum strength properties and antioxidants are required only for maximum resistance to heat In addition to its use as a stock in itself Hypalon can be blended with other types of rubbers to provide a wide range of properties

Extreme resistance to ozone is the most important property of Hypalon Its chemical resistance to strong chemicals such as nitric acid sulfuric acid, chromic acid hydrogen peroxide, and strong bleaching agents is superior to any of the commonly used rubbers These vulcanizates also have good heat resistance, mechanical properties and unlimited colorability

Typical applications include covers for conveyor belts, steam hose tubes, O rings and gaskets in ozone generators, miscellaneous molded goods, and coated fabrics for outdoor service

Polyurethane rubbers. Polyurethane rubber is an all inclusive name given to elastic polymers containing the urethane linkage In general they are prepared by extending a polyester, a polyalkylene ether glycol, or a polyhydrocarbon diol with diisocyanates, the extension of bis chloroformates of such polymeric diols with diamines is also a practical method of rubber synthesis In either case the chosen diol is usually in the 1000–5000 molecular weight range

The most common polyesters are adipates of C_4, C_3, or C_2 glycols, whereas polybutylene ether glycol from the cationic polymerization of tetrahydrofuran dominates the polyether field Little has been reported on the hydrocarbon diols with exception of polychloroprene diols

A basic repeating structure for a typical urethane rubber can be graphically depicted as shown below

The cross linking reactions for converting polyurethane rubbers to useful products depend upon the particular type of processing steps desired The so called millable or processible rubbers can first be fully compounded and then subsequently milled extruded calendered, and molded much like the diene rubbers The actual vulcanization reactions are commonly conducted at temperatures below 300°F Cross linking was first achieved by diisocyanate reactions with active hydrogens in the polyurethane rubber chain Current practice is to prepare the polyurethane chain with reactive centers such as olefin bonds so that the more conventional vulcanization by sulfur is possible

The casting polyurethane rubbers are processed differently from the millable rubber and are distinctively characterized by a relatively short pot life after being fully compounded In a typical case a polyester is pretreated with a molecular excess of di or polyisocyanate The resulting prepolymer in a molten condition is mixed with a reticulating agent such as an aromatic diamine and quickly poured into a mold Chain extension and cross linking occur concurrently in casting systems

$$\left\{ O{-}(CH_2)_4{-}[O{-}(CH_2)_4]_m{-}O(CH_2)_4{-}OC(=O){-}NH{-}C_6H_3(CH_3){-}NH{-}C(=O) \right\}_n$$

Mixing This step accomplishes an intimate and homogeneous mix of the ingredients required by the recipe. It is carried out either on a *two-roll mill* (Fig 1), or in an internal mixer Virtually all stocks mixed in large volume are now mixed in internal mixers

The internal mixing procedure involves the addition of the ingredients through the loading chute in the proper order and with suitable intervals of mixing between each addition For the recipe given above the mixing time is about 8 minutes.

There are thousands of different recipes each designed for different purposes Each of these requires a mixing procedure of its own In addition, various manufacturers differ in their ideas as to precisely how the mixing operation should be conducted and the equipment in various plants differs widely The trend in the industry is toward more automation in this process, both in weighing and charging the ingredients into the mixer and in handling the stock after discharge.

Forming. Usually forming operations involve either extrusion into the desired shape or calendering to sheet the material to some specified gage or to apply a sheet of the material to a fabric Figure 2 shows a calender In the case of tires the tread and sidewall may be formed by extrusion to give a shape as illustrated in Fig 3 The cord reinforcing plies are coated with rubber on a calender

Building Building operations varying from simple to complex are required for products such as tires, shoes fuel cells press rolls conveyor belts and life rafts These products may be built by combining stocks of different compositions or by combining rubber stocks with other materials of construction such as textile cords woven fabric or metal For a few products no building operations are required For example many molded products are prepared for molding by extrusion and cutting into lengths of the proper weight to place in the mold cavities Also extruded products which are given their final shape by the die used in the extruder are ready for the final step of vulcanization without any intermediate building operations



Fig 1 A two-roll mill (B F Goodrich Co)



Fig 2 A calender (B F Goodrich Co)

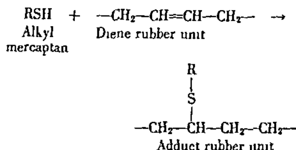
In building tires the "green" tire is built on a rotatable collapsible drum, the diameter of which is slightly larger than the diameter of the bead rings A ply of coated cord fabric cut on a bias is applied to the drum so that the cords form an angle of about 40° with a circumferential line around the drum A second ply is then applied with the cords running at about 90° to those in the first ply The bead assemblies are then applied and the edges of the first two plies are folded over the beads A third and then a fourth ply are added with alternation of the direction of the cords and their edges are folded over the beads The extruded tread strip which has previously been cut to the proper length and spliced is then applied and rolled down The drum is then collapsed and the green tire removed

... section of a ... Following ... "lifted" by inserting a rubber curing bag This operation increases the diameter of the green tire in the crown or tread area and draws the two beads together, thus forming a section resembling that of the finished tire Sometimes this operation is carried out automatically in the curing press

The building operations for tires vary widely, depending on the kind of tire to be built and the equipment available in the particular plant For example the cord reinforcement may be of rayon nylon or steel wire, the number of plies required increases as the size and service requirements increase, the tread proper may be of a different composition than the tread base and sidewalls, one of the sidewalls may be made of a white stock, the construction may be of the tubeless variety in which case an air barrier ply of rubber is ...

... prod ... tion ... in the event that

synthetic rubbers Although the preparation of these rubbers covers a complex set of chain reac tions involving several types of double bonds the basic reaction involves a low molecular weight alkyl mercaptan and a conventionally prepared diene rubber



The reaction is usually carried out in emulsion form and the general technology and equipment used in making SBR rubbers can be adapted to the preparation of the adduct derivatives Variations in the nature of the base polymer the mercaptan used and the extent of saturation give a wide range of compositions and physical properties

The technology of using adduct rubbers such as processing compounding techniques curing reci pes and fabrication methods is similar to the meth ods conventionally used for the standard diene rubbers

Interesting commercial possibilities are offered by the products having a high degree of satura tion Adduct rubbers present competition in the field of specialty rubbers because of the good aging, heat resistance low temperature properties solvent resistance low gas permeability or ozone resist ance in one or another of the derivatives Outstand ing performance in these qualities has been achieved by adducts of polymers whose double bonds had been over 90% saturated

Typical end use applications include those fields in which specialty rubbers have proved necessary and preferable They have exhibited outstanding performance when used as removable white side walls for tires radiator coolant hose impregnant for airship fabric and oil resistant foam for rail road journal boxes See POLYMER POLYMERIZA TION RUBBER PRODUCTS [R P D]

Bibliography H L Fisher *Chemistry of Natural and Synthetic Rubbers* 1957 P J Flory *Princi ples of Polymer Chemistry* 1953 J Le Bras *Rub ber Fundamentals of Its Science and Technology* 1957 G S Whitby (ed) *Synthetic Rubber* 1954

Rubber products

Articles in which a rubber is the most essential ingredient Because of the wide variety of rubber products many of which are made by methods developed specifically for them this article will be limited to the more common methods of manu facture adapted to the widest variety of products

With a few minor exceptions the manufacture of rubber products involves the following processes (1) compounding (2) mixing (3) forming (4) building or assembly and (5) vulcanization

Compounding The process of weighing the var ious ingredients required by the formulation is called compounding The term is also used for the technical problem of determining what the formula tion should be This involves the selection of the rubber the rubber chemicals (accelerators and antioxidants) the reinforcing pigments or fillers and the processing aids all of which must be care fully selected and proportioned with the objective of making a product economically and with the requisite properties to perform its function in a satisfactory manner After the formulation has been determined it is written as a recipe from which the weighing step proceeds A typical recipe for a passenger car tire tread is shown in the table

Recipe for passenger-car tire tread rubber

Ingredients	Base recipe parts	Recipe for 1000 lb batch lb
SBR 1502*	100	607
HAF carbon black†	50	303 5
Processing oil	8	48 5
Zinc oxide	3	18 2
Sulfur	1 75	10 6
Accelerator	1 0	6 1
Antioxidant	1 0	6 1

* A butadiene-styrene copolymer polymerized at 5°C (a cold rubber)
† HAF high-abrasion furnace

The base rubber a cold SBR is currently the most widely used rubber for passenger car tires Some manufacturers prefer an oil extended SBR as the base polymer This differs from the cold SBR in being made to a considerably higher viscosity and is extended at the latex stage by the addition of 25 or more parts of a petroleum oil to each 100 parts of polymer When the oil is used to facilitate processing the proportion of carbon black is cal culated on the combined oil and rubber present Thus if in the above recipe a 37 5 part oil extended rubber were used the proportion of carbon black on the combined oil and rubber that is 137 5 parts would be 68 75 parts

The carbon black is an essential and critical com ponent Its use facilitates processing but more importantly it confers excellent wearing character istics The proportion used varies between about 45 and 55 parts See CARBON BLACK

The sulfur accelerator and zinc oxide represent the vulcanization system The sulfur is the vulcan izing agent and the accelerator and zinc oxide to gether with the fatty acid already present in the SBR provide the required acceleration of the cure The kind of accelerator used varies from one manu facturer to another but its selection is based on providing the maximum safety in processing with a fast rate of cure

An antioxidant is added to confer some protec tion against tread cracking The base rubber con tains a stabilizer which is often a good antioxidant

The quantities shown in the right hand column of the table are weighed accurately and placed in suitable containers for charging into the mixer

See GERANIALES, RUBBER

Rubellite

The red to red violet variety of the gem mineral tourmaline Perhaps the most sought for of the many colors in which tourmaline occurs, it was named for its resemblance to ruby The color is thought to be caused by the presence of lithium Fine gem quality material is found in Brazil, Madagascar, Maine, Southern California, the Ural Moun

hardness of 7-7½ on Mohs scale, a specific gravity near 3.04, and refractive indices of 1.624 and 1.644 See GEM, TOURMALINE [R T L.]

Rubiales

quinine), gardenia bluets coffee ipecac (used medicinally) and madder (grown for its red dye) Elder viburnum, honeysuckle (*Lonicera*), and coralberry are members of the honeysuckle family (Caprifoliaceae) The valerian family (Valerianaceae) yields the medicinal spikenard, and teasel of the Dipsacaceae provides the dried fruit heads used by the fuller to raise the nap on cloth See CINCHONA COFFEE, IPECAC, QUININE, see also DICOTYLEDONEAE, EMBRYOPHYTA PLANT KINGDOM [P D S]

Rubidium

With an atomic number of 37 and an atomic weight of 85.48 rubidium Rb stands near the bottom of the alkali metal family in group IA in the periodic

table of the elements It is a soft, light, low melting reactive metal Relatively little is known about rubidium because it has never been available in quantity at a reasonable price In 1958, rubidium salts became much more readily available and at lower prices as by products of lithium chemicals manufacture, and knowledge of the properties and reactions of rubidium metal should develop accordingly Rubidium was discovered in 1861 by R Bunsen and G. R. Kirchhoff by interpretation of its spectral lines. Bunsen prepared free rubidium metal for the first time that same year using an electrolytic method

Uses Most uses of rubidium metal and rubid-

in treating goiter and syphilis. Rubidium mercury amalgams have been used as catalytic agents

Occurrence. Rubidium is a fairly abundant element in the earth's crust, being present to the extent of 310 parts per million (ppm). This places it just below carbon and chlorine, and just above fluorine and strontium in abundance—well ahead of chromium, zinc, nickel, copper, and lithium Sea water contains 0.2 ppm of rubidium, which, although low, is twice the concentration of lithium Traces of rubidium are found in sea water plants and animal organisms

Rubidium is like lithium and cesium in that it is tied up in complex minerals, it is not available in nature as simple halide salts as are sodium and potassium The major source of rubidium is lepidolite which may contain up to 3% Rb_2O. Carnallite and pollucite are other minerals in which rubidium is found These various minerals are found in the United States in California, South Dakota, New Mexico, and Maine Rubidium containing lepidolite is found in substantial quantities in Rhodesia as well

Metallurgical extraction. Rubidium metal is not produced on a commercial scale In the limestone process for the conversion of lepidolite ore to lithium chemicals, however, a mixed alkali carbonate liquor is obtained by carbonation in a submerged combustion evaporator after the separation of the bulk of the lithium values as the hydroxide. Filtration after carbonation removes lithium carbonate and gives a filtrate containing carbonates of potassium, rubidium, and cesium

The separation of the mixed alkali salts may be effected by a variety of methods but the most common scheme involves complex formation. The order of ease of complex formation in the alkali metal series is Cs>Rb>K>Na>Li. Stannic chloride and zinc ferrocyanide are among the most economical compounds which form insoluble complexes with rubidium Actually, the bulk of the potassium is first removed by precipitation as the bicarbonate before proceeding to the separation of rubidium and cesium When stannic chloride is used, the rubidium and cesium chlorostannates are precipitated separately (the cesium salt is more insoluble and comes out first), and are then decomposed to the chlorides with recovery of the stannic chloride Sodium zinc ferrocyanide may be employed in a similar separation, precipitating first

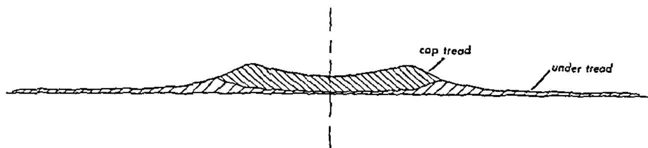


Fig 3 An extruded tread and sidewall section

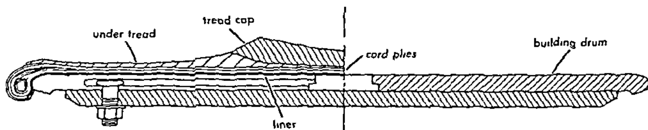


Fig 4 Cross section of building drum and green tire

no building operations are involved the forming operation This is the process which converts the essentially plastic raw mixture to an elastic state It is normally accomplished by applying heat for a specified time at the desired level The most common methods for vulcanization are carried out in molds held closed by hydraulic presses and heated by contact with steam heated platens which are a part of the press, in open steam in an autoclave under water maintained at a pressure higher than that of saturated steam at the desired temperature in air chambers in which hot air is circulated over the product, or by various combinations of these

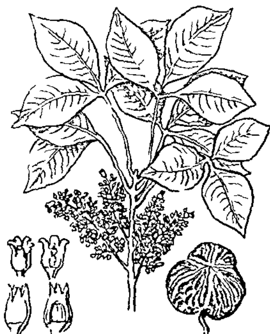
The time and temperature required for vulcanization of a particular product may be varied over a wide range by proper selection of the vulcanizing system The usual practice is to use as fast a system as can be tolerated by the processing steps through which the material will pass without "scorching," that is without premature vulcanization due to heat during these processing steps Rapid vulcanization effects economies by producing the largest volume of goods possible from the available equipment This is particularly the case for products made in molds because molds are costly and the output from them is determined by the number of heats which can be made per day

The rate of vulcanization increases exponentially with an increase in temperature and hence the tendency is to vulcanize at the highest temperature possible In practice this is limited by many factors, and the practical curing temperature range is 260–340°F There are numerous exceptions both below and above this range, but it probably covers 95% of the products made See RUBBER

Finishing operations following vulcanization include removal of mold flash, sometimes cutting or punching to size, cleaning, inspection for defects, addition of fittings such as valves or couplings, painting or varnishing, and packing {A SU}

Rubber tree

The plant *Hevea brasiliensis* a member of the spurge family (Euphorbiaceae) and a native of the Amazon valley It is the natural source of commercial rubber The tree may become 60–100 ft tall and



Foliage and flowers of the rubber tree (*Hevea brasiliensis*) (From P DeJonville, *Atlas de Poche des Plantes Utiles des Pays Chauds, Librairie des Sciences Naturelles, 1902*)

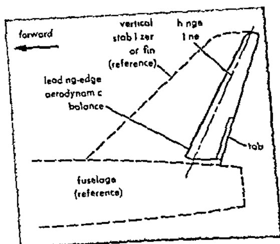
8–10 ft in circumference It has been introduced into all the tropical countries supporting the rain forest type of vegetation, and is grown extensively in established plantations especially in Malaysia The latex from the trees is collected and coagulated The coagulated latex is treated in different ways to produce the kind of rubber desired Rubber

been likened to the color of pigeon's blood Star rubies do not command comparable prices but they, too are in great demand *See* GEM [R T L.]

Rudder, aircraft

The hinged rear portion of an aircraft vertical stabilizing surface, used to obtain directional or yaw control moments. The angular setting of the rudder is controlled by the human or automatic pilot through the flight control system A typical rudder control surface includes aerodynamic balance and tab features as illustrated The principles of operation of rudder control surfaces and flutter prevention methods are identical with those for the elevator and for aileron control surfaces

Flight maneuvers Changes in aircraft heading are accomplished by banked turns, much as highway curves are banked to avoid lateral acceleration on automobiles in high speed turns The aircraft's lateral controls or ailerons are the banking controls it is the rudder s primary function to coordinate the bank and prevent yaw caused by rolling Other functions of the rudder are the maintenance of straight ground paths during take off and landing and the recovery from stalls and spins On guided missiles with cruciform wings changes in heading may be accomplished by yawing rather than banking and the rudder becomes the primary turning control

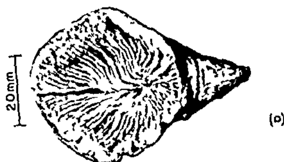


Typical rudder surface

Special applications All moving vertical stabilizers replace rudders on some supersonic aircraft just as all moving horizontal stabilizers replace elevator control surfaces The rudder may be geared to move with the ailerons (two-control) or combined with the elevator surfaces when the horizontal and vertical stabilizers are combined in a V tail *See* FLIGHT CONTROLS [M J AB]

Rugosa

An extinct order of corals which flourished during the Paleozoic Era The Rugosa or Tetracorallia, first appeared in the Ordovician and became extinct in the *Permian* Nothing is known of the soft parts Their morphology and classification are based on the skeletal structures Both simple and compound skeletons are common, the simple ones having typically a curved horn shaped appearance, and the compound ones forming groups of cylindrical stems, or polygonal columns The simple form is called a corallite and the compound one a corallum At the top of each corallite is a cuplike depression called the calyx







Rugose corals (a) Simple, horn-shaped corallite, showing calyx with septa and cardinal fossula (b) Cylindrical corallum with corallites having calyxes *each* with an axial aulos (c) Polygonal corallites forming massive corallum

The *internal* skeletal structures consist mainly of three elements the septa tabulae and dissepiments The septa are vertical plates arranged with

Phys cal properties of rubid um metal

Property	Temperature °C	Temperature °F	Metric (scientific) units	British (engineering) units
Density	20	68	1 53 g/cm³	95 5 lb/ft³
Melting point	39	102		
Boiling point	688	1270		
Heat of fusion	39	102	6 1 cal/g	10 95 Btu/lb
Heat of vaporization	688	1270	212 cal/g	381 Btu/lb
Viscosity	50	122	6 26 millipoises	4 1 kinetic units
	220	428	3 23 millipoises	
Vapor pressure	294	561	1 mm	0 019 lb/in²
	628	1162	400 mm	7 75 lb/in²
Thermal conductivity	39	102	0 07 cal/(sec)(cm²)(cm)(°C)	16 9 Btu/(hr)(ft²)(°F)
	50	122	0 075 cal/(sec)(cm²)(cm)(°C)	18 1 Btu/(hr)(ft²)(°F)
Heat capacity	39 126	102 259	0 0913 cal/(g)(°C)	0 0913 Btu/(lb)(°F)
Electrical resistivity	50	122	23 15 microhm-cm	
	100	212	27 47 microhm-cm	

cesium then rubidium These precipitated ferrocyanides can be decomposed thermally to the carbonates

Rubidium metal can be made by electrolysis of the chloride or by thermochemical reduction of the carbonate with a metal such as magnesium or with a reducing compound such as calcium carbide Magnesium reduction has been employed in many of the small scale operations which have been practiced to date

Physical properties The physical properties of rubidium metal are summarized in the table

Chemical properties Rubidium is so reactive with oxygen that it will ignite spontaneously in pure oxygen The metal tarnishes very rapidly in air to form an oxide coating and it may ignite The oxides formed are a mixture of Rb_2O Rb_2O_2 and RbO_2 The molten metal is spontaneously flammable in air Rubidium also is reputed to form an ozonide

Rubidium reacts violently with water or ice at temperatures down to $-100°C$ It reacts with hydrogen to form a hydride which is one of the least stable of the alkali hydrides Rubidium does not react with nitrogen With bromine or chlorine rubidium reacts vigorously with flame formation Rubidium dissolves in liquid ammonia in the presence of metallic catalysts or with gaseous ammonia in the absence of catalysts to give rubidium amide $RbNH_2$ and hydrogen Rubidium reacts with carbon monoxide to give the carbonyl RbCO

Organorubidium compounds can be prepared by techniques similar to those used for sodium and potassium In general the behavior of rubidium in organic reactions is probably similar to that of sodium and potassium but specific data are unavailable due to the cost and unavailability of rubidium metal This has hindered research on its reactions For a discussion of handling techniques see SODIUM

Availability In 1958 rubidium metal was sold only in gram lots in sealed ampules Total production was probably less than 1 lb/year and the price was about $4.50 per gram (about $2000 per lb) The cost of rubidium salts had been reduced from several hundred dollars to $15–30 per lb by 1958

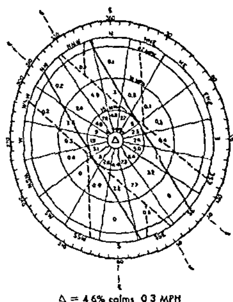
Principal compounds There are really no principal rubidium compounds in the sense of important uses or volume of production

Analytical methods Rubidium may be detected qualitatively by the red color obtained when volatile salts are introduced into a gas burner flame Separation of rubidium compounds from those of the other alkali metals is difficult but can be accomplished by ion exchange. Quantitative determination can be made by flame photometry or by the use of one of various precipitants, such as tetraphenylboron derivatives or the uranyl acetates See ALKALI METALS CESIUM [M SI]

Ruby

The red variety of the mineral corundum in its finest quality the most valuable of gem stones. Only medium to dark tones of red to slightly violetish red or very slightly orangy red are called ruby light reds purples and other colors are properly called sapphires In its pure form the mineral corundum is colorless The rich red of fine quality ruby is the result of the presence of a minute amount of chromic oxide usually well under 1% See CORUNDUM SAPPHIRE

The mineral corundum is commonly a constituent of basic igneous rocks but it rarely occurs in a transparent form suitable for gem use It is also known as a constituent of the type of marble formed in the contact zone of an igneous intrusion into an impure limestone It is in this type of deposit that the finest rubies (those of Mogok Burma) were formed Today most of those mined in the Mogok region are taken from the famous gem gravels of that area There are only two other sources of significance Ceylon and Thailand In each of these countries the finest quality obtained is far less valuable than fine Burma material for the Ceylon ruby is too light in color and the so-called Siam ruby is a darker red The finest ruby is the transparent type with a medium tone and a high intensity of slightly violet red which has



Δ = 4 6% calms 0 3 MPH

Fig 2 Wind rose Figures indicate percentage of directional winds in each velocity group

single runway pattern is approximately 40 operations (landings or take offs) per hour

Runway orientation Analysis of low visibility and all weather wind data tabulations, obtained from the United States Weather Bureau indicates the best orientation of single or parallel runways and may also dictate the use of cross-wind runways The Federal Aviation Agency (FAA) prescribes that aircraft should land at least 95% of the time with cross-wind components not exceeding 15 mph The FAA has replaced the former Civil Aeronautics Administration (CAA)

The maximum single runway wind coverage percentage may be graphically determined by plotting percentages of directional winds in the 4 15 31 and 47 mph velocity groups on a wind rose base as shown in Fig 2 A pair of parallel lines representing a runway is then superimposed tangent to the 15 mph velocity scale and oriented in such a manner that the total of the percentages included between the lines is a maximum The runway oriented 150–330° (solid lines) will provide 95% wind coverage with an allowable cross wind component not exceeding 15 mph Two runways oriented 130–310° and 180–360° (dotted lines) would provide 99% coverage

Although maximum wind coverage is operationally desirable it is even more important to obtain obstruction free approach zones and to avoid air corridors and heavily populated areas These objectives together with topographical considerations will often require a compromise in runway orientation The alignment of runways with existing directional radio range facilities is also desirable because a straight in approach is then permitted the ceiling and visibility minimums are reduced, and the operational capabilities of the field are improved

Runway length and loading. The four FAA classes of air carrier service—local, trunk, continental and intercontinental—have been assigned maximum effective runway lengths of 4200, 6000, 7500 and 10,500 ft and maximum equivalent single-wheel pavement loadings of 30 000, 60 000, 75 000, and 100 000 lb, respectively

The runway length and strength for a particular airfield is determined, within the appropriate class of service by the route pattern and the type of aircraft to be flown The indicated maximum effective runway lengths are based on standard 59°F temperature plus 41°F, at sea level elevation, and zero runway gradient Runway lengths are increased for higher airport elevation at the rate of 7% for each 1000 ft above sea level and for runway gradient at the rate of 20% for each 1% of effective runway gradient (the maximum difference in runway elevation divided by the total runway length)

As a minimum requirement the longitudinal centerline profile of a runway is designed to provide an unobstructed line of sight from any point 5 ft above the runway to any other point 5 ft above the runway and within a distance of at least 500 ft plus one-half the ultimate runway length

Runway markings. FAA standards for runway marking include magnetic azimuth of the runway,

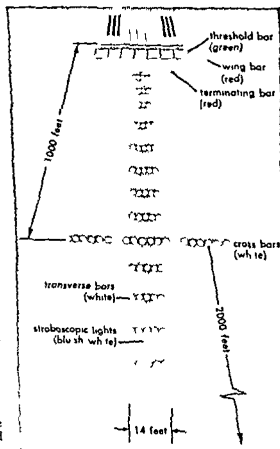


Fig 3 Standard FAA approach lighting

tetrameral symmetry They are visible on the floor and walls of the calyx The first four septa to appear are called the protosepta and are known individually as the cardinal, counter, and alar (lateral) septa When the corallite is oriented with the cardinal septum farthest away, it can be seen that all later septa are arranged pinnate to the cardinal septum and parallel to the counter septum This results in the septa being arranged in four quadrants, known as the cardinal and counter quadrants All later septa commonly are alternately long and short, and are called major and minor septa, respectively.

The protosepta may undergo various types of modifications In many forms the cardinal septum becomes shortened, leaving a keyhole slot called the cardinal fossula Counter and alar fossulae are known in some forms Other septal modifications include carinae, lateral ridges on each septum, appearing as series of cross bars, acanthine septa, composed of vertical rows of spines, and degenerate septa appearing as discontinuous ridges in the calyx

An axial structure, the columella is present in some forms This is an axial rod of calcite produced by an enlargement of the counter septum or as an independent structure In most forms with a columella, the ends of the major septa are attached to it In other forms, the axial ends of the septa are elevated to form a boss, or laterally deflected to form a central tube called an aulos

The tabulae are horizontal or arched plates extending between the walls of the corallites The dissepiments are cystose structures lying between the outer parts of the septa in some forms The dissepiments replace the septa in some genera See COELENTERATA FOSSILS [E C ST]

Running fit

A running fit permits mating parts to move relative to each other A free running fit has liberal allowance, it is used on high speed rotating journals or shafts A medium fit has less allowance it is used on low speed rotating shafts and for sliding parts Running fits are affected markedly by their surface finish and the effectiveness of lubrication See ALLOWANCE

Running and sliding fits are standardized into nine classes Close sliding fits accurately locate parts with some sacrifice in free motion, they permit no perceptible play Sliding fits permit the parts to move but are not intended for freely running parts or moving parts subject to appreciable temperature change Precision running fits permit parts to run freely at low speeds and at light journal pressures provided temperature differences are limited Close, medium, and free running fits are intended for progressively higher surface speeds journal pressures, and temperature ranges Loose running fits are for use with cold rolled shafting and tubing made to commercial tolerances

[P H B]

Runway, airport

The paved or turfed strips provided for landing and take-off of aircraft The use of turfed strips is normally limited to small aircraft with light wheel loads Paved runways are either of bituminous concrete or portland cement concrete They vary considerably as to number, orientation, length gradient, and strength, depending upon the type and number of aircraft to be accommodated and upon the elevation, climate, and topography at the airport See AIRPORT ENGINEERING

The required number of runways is determined in part by air traffic density forecasts The critical periods occur during instrument flight rule (IFR) weather and when aircraft arrivals and departures reach a peak, usually from late afternoon to early evening The average IFR capacity of several typical runway configurations is shown in Fig 1 For purposes of comparison, the capacity during visual flight rule (VFR) or contact weather of a

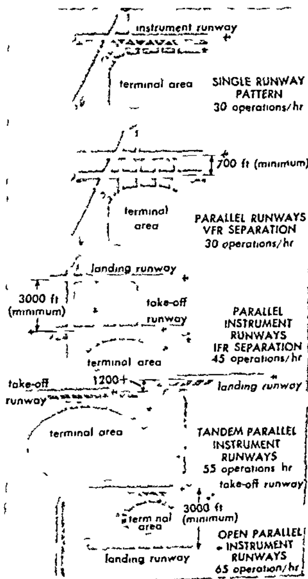


Fig 1 Typical runway configurations and average IFR capacity

This acceptance of electric energy is largely the result of its economic and labor saving advantages because it has been established that for any task a 1 horsepower electric motor can perform a farmer will require 10 times as much time to accomplish it manually Electric power is also more versatile and more dependable than other kinds of power

Dairy farm applications of electrical energy include lighting pumping water milking cows cooling milk, heating water for sanitation electric fencing ventilating and cleaning the stable elevating and drying hay feeding ensilage and handling milk in bulk through pipelines from cows to tank trucks (Fig 2)

Poultry farmers use electricity for lighting the flock to stimulate winter egg production incubating eggs brooding chicks automatic feeding and water

mechanical
onmental control

The ncreasing use of electric energy on farms

most unlimited *See* ELECTRIC POWER SYSTEMS [C N T]

Bibliography U F Earp *Rural Electrification Engineering* 1956 C N Turner (ed) *Farm Electrical Equipment Handbook*

Rural sanitation

Those procedures employed in areas outside incorporated cities and not governed by city ordinances that act on the human environment for the purpose of maintaining or improving public health The purpose of these procedures is the furtherance

1900

has helped to prolong human life more than any other single public health measure Organisms which produce such diseases as typhoid dysentery and cholera may survive for a long time in polluted water and prevention of contamination of water supplies is imperative to keep down the spread of such diseases Water tight *covers for* wells are important means of preventing surface contamination to the water supply in rural areas

Purification of a surface water supply such as that from a lake or a farm pond is accomplished by means of sedimentation filtering and chlorination Sedimentation can be effected in a storage chamber by the addition of aluminum sulphate which flocculates the finer particles of soil and ter in suspension in the water Filtering through fine sand removes the flocculated particles. Finally, the water is purified by means of a chlorine solution at the rate of ½-1 part of chlorine to 1 000 000 parts of water A system of treatment for farm pond water is diagrammed in Fig 1

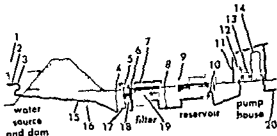


Fig 1 Farm pond water treatment system

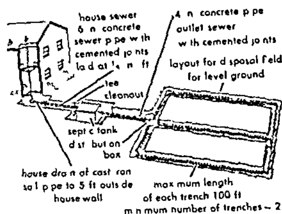


Fig 2 Typ cal fam ly s ze sewage-d sposal system

The colon bacillus is the usual indicator of pollution of water supplies by human waste Chlorine will kill such organisms and therefore it is widely used for purification of water supplies

Sewage disposal The *problem* of safe disposal of sewage becomes more complex as *population* increases The old practice of piping sewage to the *nearest* body of water has proved to be dangerous Sanitary engineering techniques are now being used in *rural areas* as well as in cities for the disposal of household and human wastes Where sewage-plant facilities are not available the most satisfactory method of sewage disposal is by means of the septic tank system as shown in Fig 2

The septic tank system makes use of a water tight tank for the receiving of all sewage Bacterial action takes place in the septic tank and most of

measured from the approach end, painted in 60-ft high numerals near each end of the runway (numerals are in tens of degrees), runway centerline stripes, 120 ft long and 200 ft on center, runway distance markers 500 ft apart indicating the final 2000 ft of runway, and continuous runway edge stripes, which serve to outline the usable runway pavement The runway markings are painted directly on the pavement with reflective white paint

Airfield lighting Airfield lighting is mandatory for night or restricted visibility operations Taxiways are provided with elevated blue marker lights spaced at 200 ft intervals along each edge Runways are provided with clear elevated marker lights, similarly spaced Marker lights along the final 1500 ft of the runway have yellow lenses on the side facing away from the runway's end Figure 3 shows the standard approach lighting method, as adopted by the FAA

There is a trend toward flush type light units, which are installed in the pavement and are designed to support the heaviest wheel loads Because these units offer no obstruction to aircraft, they can be used on the runway centerline near the runway end and as threshold bars and centerline approach light bars The stroboscopic light units indicated in Fig 3, create an extremely brilliant series of flashing lights which appear to streak toward the runway threshold Red obstruction light units are used to outline natural or man made obstructions hazardous to aircraft operations [E E S , R A F R]

Rural electrification

The provision and use of electricity in nonurban areas The almost universal availability of and use of electricity throughout rural parts of the United States is the result of a system of generating and distributing current from a nation wide network of central power stations by way of cross country transmission lines and rural line extensions that now reach 95% of the farms in the United States (Fig 1) Equally important to the rural population has been the development of hundreds of uses for electricity for light heat and power

Development *Use of electric energy in food* production started in California in the early 1880's

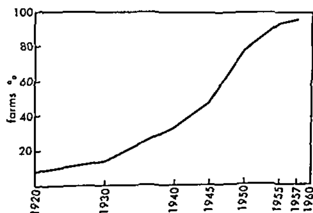


Fig 1 Per cent of farm dwellings in the United States lighted by electricity (U S Bureau of the Census)

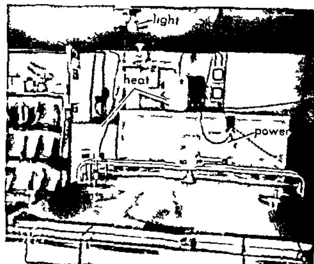


Fig 2 Electricity provides light, heat, and power in the milk house on a dairy farm. *Light is used for cleaning and sanitary care of milk utensils, heat is used for water and warming the room and electric motors provide power for pumping and cooling the milk in a bulk tank (Niagara Mohawk Power Corp)*

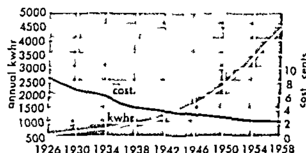


Fig 3 Trends in annual kilowatt hour use and cost of electric energy on farms in the United States where little or no irrigation is involved (*Edison Electric Institute*)

when pioneering farmers began using current from small water power driven generating plants for lighting and for pumping water for irrigation During the next 50 years a number of experimental projects were undertaken by both private and public agencies *Collectively these investigations demonstrated* convincingly the economic feasibility of using electricity widely on farms As a result national interest developed in the extension of electric lines to all rural areas In 1935 the Rural Electrification Administration (REA) was created as one of the approved projects under the Federal Emergency Relief Appropriations Act The REA was authorized to make loans for the generation transmission and distribution of electricity to rural people The farmers' cooperative was the agency through which most of the loans were made for rural line construction over the country Private projects and publicly financed ones collectively provided electric service to about 2 400 000 farms by the end of 1942

Uses More than 300 known uses of electricity have been developed for farm operations and about 150 uses are found in agricultural production

Control Control of rusts by dusting with fungicides is practical for ornamentals but breeding resistant varieties of hosts offers the greatest promise for control of all rusts Eradication of alternate hosts is effective in controlling white pine blister rust Eradication of rust in barberries is less effective, because uredospores of *Puccinia graminis* are blown great distances, but it is important, because it impedes the evolution of new races of rust See UREDINALES [R M P]

Bibliography A Stefferud (ed) *Plant Diseases*, USDA Yearbook Agr 1953 J C Walker, *Plant Pathology*, 2d ed, 1957

Rutabaga

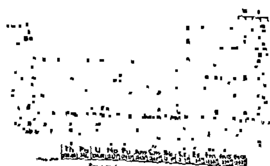
The plant *Brassica napobrassica* is a cool season hardy biennial crucifer of European origin belonging to the order Papaverales and probably resulting from crossing cabbage and turnip (see BIENNIAL PLANTS CABBAGE) Unlike turnip it has smooth nonhairy leaves and 38 chromosomes Propagation is by seed commonly sown in early summer See SEED (BOTANY) Popular yellow fleshed varieties are Laurentian and American Purple Top a leading white variety is Macomber Rutabagas have a high requirement for boron (see BORON) ... cause misshapen root growth

...

begins after frost and when the ... in diameter commonly 90–100 days after planting For diseases of rutabaga see TURNIP see also PAPAVERALES VEGETABLE GROWING [H J C]

Ruthenium

Chemical element number 44 ruthenium Ru is a hard white metal It is workable only at high temperatures and then with difficulty It is less workable than iridium but more workable than osmium



Uses Uses of pure ruthenium metal are minor Ruthenium catalysts have been used in certain organic syntheses Complex ruthenium alloys with and without added osmium are very hard and are used for pen tips nonmagnetic instrument pivots and similar items Ruthenium is added as a hardening agent to platinum and palladium used in electrical contacts and jewelry

Properties of ruthenium

Atomic weight	101 1
Crystal structure	close-packed hexagonal $a = 2.70$ $c = 4.28$ at 20 C
Density (at 20°C)	12 32 g/cm³
Melting point	2250°C
Linear thermal expansion coefficient (per °C at 20°C)	9.6×10^{-6}
Specific heat (at 0°C)	0 057 cal/°C
Electrical resistivity (at 20°C)	7 2 μ ohm-cm
Modulus of elasticity	60×10^6 psi

Metallurgical extraction. A fusion with sodium peroxide of the residue from previous refining operations which have removed platinum palladium and rhodium yields water and acid soluble compounds of iridium ruthenium, and osmium The last two elements may be removed from solution by distillation of their tetroxides chlorine is streamed through the hot solution and the resultant gas is passed through hydrochloric acid which reduces the ruthenium to the trichloride Osmium can then be distilled from the latter solution Alternatively hydrated ruthenium dioxide may be precipitated from the original fusion extract using an alcohol treatment

Physical and chemical properties. Ruthenium is not very ductile when cold although pure single crystals can be bent easily The metal can be melted by torch electric arc or electron beam and worked hot usually above 1500°C. In this manner it may be worked down to a relatively thin sheet

Ruthenium is resistant to the common acids in

...

hot attack it slightly at room temperature It is attacked rapidly by sodium peroxide but is resistant to a number of molten salts Ruthenium may be electrodeposited from a molten salt bath

Ruthenium oxidizes slowly in air above 450°C to form ruthenium dioxide which is only slightly volatile

Ruthenium compounds. Ruthenium exhibits the positive oxidation states of 2 3 4 6 7 and 8 Potassium ruthenate $KRuO_4 H_2O$ is soluble in water and ...

...

... in this manner and subsequently reduced to pure metal by heating in hydrogen Ruthenium tetroxide RuO_4 may be formed by treatment of the metal with a very strong oxidizing agent such as sodium permanganate ozone, or chlorine, the oxide is distilled off as described under metallurgical extraction Ruthenium tetroxide is highly volatile and poisonous It melts at about 25°C and has an extrapolated boiling point of 135°C. See IRIDIUM, OSMIUM, PLATINUM [H J A]

the sewage solids decompose, are given off as gases, or go out into the drainage lines as liquid The gas and liquid are then released from the top 2 ft of soil without odor or sanitary problems The solids that do not decompose settle to the bottom of the tank where they can be easily removed and disposed of safely Such a sewage disposal system has made it possible for all farm homes and rural communities to have modern bathroom equipment and sanitary methods of sewage disposal *See* SEPTIC TANK, SEWAGE DISPOSAL, WATER TREATMENT [H E ST]

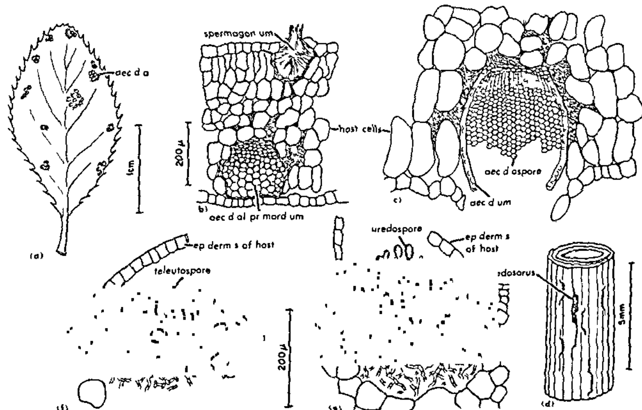
Rust (microbiology)

Fungi of the order Uredinales or diseases caused by these fungi Rusts are important plant pathogens Stem rust destroyed 60% of the wheat in Minnesota and neighboring states in 1935, flax rust caused a loss of $10 000 000 in the same area in 1951 and other rusts, such as those of oats, barley, rye white pines, apple, and snapdragon are of economic importance

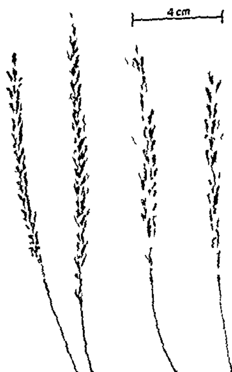
Microcyclic rusts such as *Puccinia malvacearum* (hollyhock rust) produce only teleutospores and basidiospores Macrocyclic rusts produce other types of spores in addition Autoecious rusts, such as *P asparagi* (asparagus rust) complete their life cycle on one species of host Heteroecious rusts require two species of host

Life cycle. The life cycle of *Puccinia graminis tritici*, stem rust of wheat exemplifies a macrocyclic heteroecious rust (see illustration) The haploid mycelium, which parasitizes leaves of barberry (*Berberis*), forms spermagonia and aecidial primordia Spermatia formed in spermagonia are transported by insects to compatible receptive hyphae After fusion and nuclear pairing, the aecidial primordia develop into aecidia which liberate binucleate aecidiospores Upon germination, an aecidiospore forms a binucleate mycelium capable of parasitizing wheat This mycelium produces spores of two types uredospores which spread the disease to other wheat plants, and two celled teleutospores The two nuclei in each cell fuse and the teleutospore overwinters When it germinates reduction division occurs, each cell produces an epibasidium and basidiospores are discharged Basidiospores infect barberry to complete the cycle

Specificity. Rusts vary in specificity of parasitism The white pine blister rust *Cronartium ribicola*, forms spermatia and aecidia on 11 species of 5 needled pine and the binucleate stage infects many species of current and gooseberry (*Ribes*) *Puccinia graminis* has 8 subspecies which differ only slightly in morphology but considerably in range of parasitism For example *P graminis tritici* attacks wheat barley rye and other grasses, but not oats *P graminis avenae* infects oats but not wheat barley or rye In addition over 240 morphologically identical physiological races of wheat rust are distinguishable by the pattern of infection of 12 selected varieties of wheat



Puccinia graminis tritici stem rust of wheat (a) Lower surface of barberry leaf showing aecidia (b) Cross section of barberry leaf showing spermagonium and aecidial primordium Receptive hyphae not shown (c) Section of barberry leaf showing mature aecidium (d) Portion of wheat stem showing uredosori on leaf sheath (e) Cross section of wheat leaf showing uredosorus and uredospores (f) Cross section of wheat leaf showing teleutosorus and teleutospores



Spikes of rye lemmas are short-awned and caryopses normally held loosely between lemma and palea

as important economically as other cereal grains less effort has been expended to develop superior new varieties Most of the first rye varieties are ...

... to obtain greater uniformity in grain size and color better winter hardiness less susceptibility to lodging (falling over) and higher yield One variety Tetra Petkus was developed in Germany by doubling the chromosome number of Petkus with ... ples of im ... s technique ... plants are ... generally not superior in yield or quality of grain

Cultural practices The methods used in rye production are very similar to those used for winter wheat Because of its greater winter hardiness rye can be seeded 2 to 3 weeks later than wheat When rye is used both as a fall pasture and a grain crop a common practice is to plant it earlier than winter wheat and to graze the fall growth Rye generally is seeded at the rate of 2 bushels per acre

Rye is most commonly harvested with a combine (see AGRICULTURAL MACHINERY) Because the grain is more subject to shattering than is that of other cereals one of the problems encountered in fields following rye production is volunteer growth For this reason wheat should not be planted after rye because a mixture of rye with wheat lowers its market value Rye straw is preferred over other cereal straws for bedding and consequently it is more often saved by baling following the combine See GRAIN CROPS [I J J]

Processing The selection of the proper type of grain is quite important in the manufacture of quality rye flour Plumpness soundness, and inside color are desirable characteristics Sprout damage ergot and excessively thin rye should be avoided In preparing rye for milling separators aspirators disk machines scourers and brush machines are used to remove foreign material It is desirable to remove as much of the outer covering ... milling ... neficial ... 11.5%

The *milling system* for a rye mill is similar to that for a wheat mill except that purifying and grading operations can be much simpler Rye being tougher and more starchy than wheat it requires more grinding and more bolting surface Also substantially more horsepower is required to reduce it to flour Corrugated rolls are used throughout for all breaks and reductions Smooth rolls have little value as they cause rye middlings to flatten and flake Corrugations used are usually somewhat finer than for a wheat mill with greater spiral and at least 2½ 1 differential on all grinding operations that is one of the rollers operates 2½ times faster than the other

Satisfactory white and dark rye flours can be made on a comparatively simple flow by proper selection and combining of basic mill flour streams White rye flours usually range from 0 58% to 0 64% ash with dark rye running from 2 00 to 2.50% ash Intermediate grades can be made by combinations of white and dark flours in the desired percentages Rye flour yields vary widely according to the amount of dark flour made A normal extraction ranges between 75 and 85% of the total grain ground

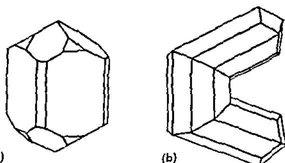
As rye flour does not contain gluten it is very difficult to bake a 100% rye loaf of bread Rye flours are usually blended with wheat flours to se...

... dlings consisting of finely ground bran germ screenings and a small amount of endosperm. See FOOD ENGINEERING.

[J A SH]

Rutile

A mineral having composition TiO_2 and crystallizing in the tetragonal system Prismatic crystals of rutile are common, they are usually vertically striated with pyramidal terminations. *Twinning is frequently observed* The hardness is 6–6 5 (Mohs scale), the specific gravity 4 2 The luster is adamantine to submetallic and the color red to reddish brown Iron is usually present in small amounts up to 10% The varieties rich in iron are black Rutile is the most common of the three mineral polymorphs of TiO_2, *the others are anatase (tetragonal)*, and brookite (orthorhombic)



Tetragonal rutile crystals (a) Prismatic with dipyramid terminations (b) Elbow twins (From C S Hurlbut, Jr, Dana s Manual of Mineralogy, 16th ed, Wiley, 1952)

Rutile is found as an accessory mineral in granites, granite pegmatites, mica schists, gneisses, and crystalline limestones and dolomites It may also be in quartz veins traversing these rocks Slender acicular crystals are often found in transparent quartz (rutilated quartz) and in phlogopite mica in which it produces asterism *Submicroscopic crystals oriented in corundum produce the star in star ruby and star sapphire*

Rutile is a common constituent of black sands associated with ilmenite, magnetite zircon and monazite from which it is recovered for industrial purposes Nearly colorless rutile produced synthetically is used as a gem stone *See* ILMENITE, TITANIUM [C S HU]

Rydberg constant

An atomic constant connected with the other universal *physical constants by the relation* $R_\infty = 2\pi^2 me^4/ch^3$ Here m and e are the rest mass and charge of the electron, c the velocity of light, and h Planck's constant. The symbol R_∞ indicates that the constant refers to a hypothetical atom of infinite nuclear mass. The values of R are experimentally determined from measurements of the wavelengths of spectral lines of actual atoms and vary slightly *with the mass of the nucleus* The deviation from R_∞ is greatest for the lightest atoms such as *hydrogen, deuterium*, and helium, the accepted values of R_H, R_D, R_{He} and R_∞ being 109,677 58, 109,707 42, 109,722 26, and 109,737 31 cm^{-1}, all ±0 01 cm^{-1} Theoretical values of R are obtained by replacing m in the preceding formula by the reduced mass of the electron and nucleus, that is, by $mA/(m + A)$, where A is the nuclear mass of the atom in question The importance of the Rydberg constant lies chiefly in the fact that it gives a precise relation between the universal constants *involved, one which must be satisfied by any* consistent set of constants that may be adopted *See* ATOMIC CONSTANTS, ATOMIC STRUCTURE AND SPECTRA [F A J]

Rye

The plant, *Secale cereale*, is grown more extensively in Europe than in America In the United States the greatest acreage is in the Great Plains states from North Dakota southward to northern Texas, either as a grain crop or for pasture. In the *eastern and southeastern states* rye is grown primarily for temporary pasture and as a green manure *crop About one-third of the rye harvested for grain* in the United States is used for livestock feed, less than one fourth is used for alcohol and other spirits, and about one-fourth is used for human food blended with wheat as a bread flour The remainder is used as seed In contrast a much higher portion of rye grown in Europe including Russia, is used for human food The average annual value of rye in the United States for the 10-year period 1945–1954 was $32 599 700

Origin and description. *Rye is generally believed* to have originated in Central Asia or in Asia Minor It is a tall growing cereal grain (Fig 1) The inflorescence is a spike with the lemmas short awned and the spikelets usually two seeded (*see* GRASS CROPS, INFLORESCENCE) The grain of all varieties is naked larger than wheat and varies in color from light tan to green The chromosome *number of all varieties is 7 Although the chromo*somes of rye are not homologous to those of wheat *these two genera can be hybridized See* BREEDING (PLANT) CHROMOSOME, GENETICS Hybrids so produced are sterile but when the chromosome number of the hybrid is doubled by using colchicine or by other means the resulting amphiploids are relatively fertile (*see* COLCHICINE) Varieties most generally grown are winter annuals seeded in the *fall to produce heads the following summer (see* ANNUAL PLANTS) Although varieties are available that can be planted in the spring and harvested the *same year these are not generally grown*

An important characteristic of winter rye is its greater winter hardiness than that of most winter wheat and its greater productivity than that of other cereal grains when grown on soils of low fertility and low water holding capacity Because rye generally is planted under these unfavorable con*ditions, its yield is lower than that of winter wheat Rye also is less subject* to attack by plant diseases and insects than are other cereal grains

Varieties. Rye is a naturally cross-pollinated crop (*see* REPRODUCTION PLANT) Therefore several of the older varieties are a mixture of many types in respect to kernel color and size, plant height, and days to *maturity Because rye is not*